IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received December 29, 2018, accepted January 13, 2019, date of publication January 24, 2019, date of current version February 14, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2894874

An Energy-Efficient and Fast Convergent Resource Allocation Algorithm in Distributed Wireless Ad Hoc Networks

WEI FENG[1], YONGXIN XU[1], XIAORONG XU[1], MINGXIONG ZHAO[2], AND YINGBIAO YAO[1]

[1]School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, 310018, China
[2]School of Software, Yunnan University, Kunming, 650091, China

Corresponding author: Yingbiao Yao (e-mail: yaoyb@hdu.edu.cn).

This work was supported in part by the National Natural Science Foundation of China under Grant 61671192, in part by the Public Welfare Plan Project of Zhejiang Province under Grant LGG19F020014, in part by the Natural Science Foundation of Zhejiang Province under Grants LY19F010011, and in part by the General Science Foundation of Zhejiang Educational Committee under Grant Y201533647.

ABSTRACT Since the initial publication of Kelly *et al.*'s seminal paper on resource allocation in wired networks, many studies based on the cross-layer design philosophy have been conducted in both wired and wireless networks. The Lagrangian duality technique has been widely adopted to solve the cross-layer optimization problem, but it is slowly convergent and sensitive to the iterative step size, especially for large networks. In this paper, a joint congestion control and power allocation second-order algorithm (JCCPA) is proposed, in which the joint optimization problem is modeled as a network utility maximization (NUM) framework, and the primal-dual interior-point method is used to solve the NUM problem. The JCCPA updates the primal variables and dual variables simultaneously according to their Newton directions; as a result,resulting in a faster convergence speed is reached. Moreover, the matrix-splitting technology is utilized to decompose the computation of the Hessian matrix and its inverse into different nodes and links so that the distributed update of source rate and node power can be implemented. The simulation results demonstrate that the JCCPA not only significantly improves energy efficiency but also has a faster convergence speed and is insensitive to step size.

INDEX TERMS Congestion control, energy efficiency, fast convergence, power allocation, wireless multi-hop ad hoc networks.

I. INTRODUCTION

With the advantages of self-organization, low cost, and easy deployment, wireless multihop ad hoc networks are widely applied to expand network coverage and increase network capacity. To improve the network performance of wireless multihop ad hoc networks, cross-layer optimization schemes for resource allocation have received a great deal of attention. In order to allocate network resources appropriately, cross-layer optimization schemes break down the traditional layering architecture and capture the dependencies and interactions among five layers [1].

In recent decades, great achievements have been made in cross-layer optimization algorithms for wireless multihop ad hoc networks [2]–[11]. Among these methods, the cross-layer optimization problem is usually modeled as a network utility maximization (NUM) framework and solved by the Lagrangian dual decomposition or subgradient methods [4]–[6]. For example, Eryilmaz and Srikant [4] investigated the joint scheduling, routing, and congestion control problem, adopted the primal-dual method to achieve fair resource allocation, and further used Lyapunov stability theory to prove the stability of the queue buffer. Zhang *et al.* [5] considered the joint congestion control and power control problem in wireless multihop networks. A power consumption cost function was introduced in the objective function, and the Lagrangian dual decomposition method was used to solve the optimization problem. Based on the physical interference model, Huang *et al.* [6] investigated the complex scene of multichannel and multiradio networks, and applied both Lagrangian dual decomposition and subgradient methods to achieve the optimization of joint congestion control, power, and channel allocation in wireless networks. These first-order approaches generally use ‘‘price’’, which is a dual multiplier, to form a feedback mechanism between the sources and

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

the links, so that the algorithms can operate distributedly with the local information. However, this kind of solution has inherent defects, such as a slow convergence rate and the difficulty of selecting an appropriate iterative step size. Therefore, the second-order algorithm based on the interior-point method has emerged as an alternative approach to these problems in recent years [7]–[11]. Zymnis *et al.* [7] considered the flow control problem based on the premise that the flow route is predefined. The primal-dual interior-point method was utilized to achieve the iterative formulas of primal and dual variables, and the iterative directions of primal and dual variables were computed using Newton's method. The algorithm has a faster convergence rate; however, it is centralized, and the signaling overhead is disastrous, especially for large, dense networks. The difficulties associated with the distributed implementation of second-order algorithms lie in taking the inverse of the Hessian matrix and other matrices in the iterative direction, which contain a wealth of global information and usually cannot be turned into solvable diagonal structures. To solve this problem, a distributed Newton method was proposed in [9], which obtains approximate iterative directions of primal and dual variables and realizes the distributed variable updates. Unfortunately, the control strategy of iterative step size in [9] is rather complex in its implementation. Liu and Sherali [10] further considered the multipath routing problem based on [9] and designed a new distributed Newton method, which used the matrix-splitting approach to decompose the computation of the Hessian matrix and its inverse into source nodes and links. However, the results are approximate. In a recent work [11], the authors developed an efficient second-order distributed algorithm based on the primal-dual interior-point method for joint congestion control and routing optimization. This algorithm considers the scenario in which each network node maintains an independent queue backlog, and it uses the Sherman-Morrison-Woodbury matrix inversion formula to decompose the Hessian matrix and its inverse. The true solutions of the matrix inversion can be obtained through less information interaction, and a faster convergence speed is reached by updating the primal and dual Newton directions simultaneously. The algorithm in [11] goes a step further than the previous second-order algorithms, in which only the primal variable is updated in Newton direction.

As the above discussion indicates, the second-order resource allocation methods have been improved in some respects, but they are still largely an open problem. All existing second-order resource allocation methods for congestion control in wireless multihop ad hoc networks are based on the premise that the capacity domain is known and fixed and that even the optimization problem itself is a linear convex optimization problem. In fact, the capacity domain is usually dynamic due to the network resource allocation, channel variation, node mobility, and so on. The second-order algorithm presented here is designed to escape the bottleneck link in the network rather than allocate more power to the bottleneck link in order to establish a larger transmission link and transmit packets faster through the bottleneck link. Based on these problems, an excellent power control method should devote itself to allocating appropriate power to requisite links in order to increase the capacity of the bottleneck link, improve energy efficiency, and achieve a faster convergence rate. There are two critical hindrances to realizing this idea: identifying a way to find the bottleneck link and a way to stabilize the flow rate fluctuation in the network as a whole, which is caused by allocating the power to one link.

This paper considers wireless ad hoc networks with node power constrained, such as wireless sensor networks, and examines the problems of joint congestion control and power allocation. An energy-efficient and fast convergent second-order algorithm based on the primal-dual interior-point method is proposed, and the completely distributed updating of the flow rate and link power is realized with the matrix-splitting method. The main differences between the proposed algorithm and the current state of the art and the most relevant contributions of this paper are as follows:

1) We propose a distributed joint congestion control and power allocation second-order algorithm (JCCPA) in wireless multihop ad hoc networks for the first time, and demonstrate its convergence. This algorithm not only overcomes the slow convergence drawbacks of conventional first-order algorithm but also greatly improves the energy efficiency of the existing second-order algorithms.
2) According to the optimization model with the dynamic channel capacity constraint, we propose a modified nonlinear system comprised of perturbed KKT (Karush-Kuhn-Tucher) conditions. Compared with the common nonlinear system in the existing second-order algorithm, the modified nonlinear system introduces a modified matrix defined by the dynamic channel capacity constraint, which is adaptive to different channel capacity models, and provides a pointcut for the studies on the second-order algorithm under the interference-limited or multichannel networks.
3) We evaluate the proposed JCCPA through extensive simulations. Simulation results demonstrate that the convergence speed of JCCPA is more than 30 times faster than that of the traditional first-order algorithm and JCCPA can improve the energy efficiency by 37.5% compared to the newest second-order algorithm.

The rest of the paper is organized as follows. In Section II, we describe the system model and formulate the optimization problem. The corresponding congestion control and power allocation policies are presented in Section III. The simulation results are discussed in Section IV, and a conclusion is provided in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. SYSTEM MODEL

Consider a wireless ad hoc network with N nodes and N logical links. Let a graph $\mathcal{G} = \{\mathcal{N}, \mathcal{L}\}$ denote it, where $|\mathcal{N}| = N$ is the set of nodes, and $|\mathcal{L}| = L$ is the set of links, respectively. Tx (l) and Rx (l) are the transmitting and receiving nodes of link l, respectively. Let S denote the end-to-end flow number and $\mathcal{F} = \{f_1, \ldots, f_s, \ldots, f_S\}$ denote the set of flow rates and the cardinality of the sets $\mathcal{F}$ is $|\mathcal{F}| = F$. For each flow s, a sequence of connected links $l \in L(s)$ forms a route originating at the source node Src(s) and destined for the destination node Dst(s). Let D denote the nodes without output link. For convenience, we define a routing matrix $\mathbf{R} \in \mathbb{R}^{L\times S}$ and a node-link matrix $\mathbf{T} \in \mathbb{R}^{(N-D)\times L}$ after removing the nodes without output link. The entries $(\mathbf{R})_{ls}$ and $(\mathbf{T})_{nl}$ are defined as follows: $(\mathbf{R})_{ls} = \begin{cases} 1, & \text{if } l \in L(s), \\ 0, & \text{otherwise.} \end{cases}$ and $(\mathbf{T})_{nl} = \begin{cases} 1, & \text{if } n = \text{Tx}(l), \\ 0, & \text{otherwise.} \end{cases}$

B. PROBLEM FORMULATION

All the channels are modeled by large-scale fading with path loss exponent α along with small-scale Rayleigh fading. The instantaneous received signal-to-noise ratio (SNR) at link l can be given as $\text{SNR}_l = \frac{p_l|h_l|^2}{d_l^\alpha}$, where p_l denotes the transmit power of link l, and d_l and h_l represent the distance and channel coefficient of link l, respectively. We assume that $|h_l|^2$ follows exponential distributions with mean equal to one [16]. For convenience, the noise power is normalized here. Therefore, the effective capacity of the logical link l can be modeled as $C_l(p_l) = B\log(1 + \frac{p_l|h_l|^2}{d_l^\alpha})$,where B is the channel bandwidth. As a result, in contrast to the aforementioned studies, data rates attainable on wireless links are not a fixed number C; instead, they are a nonlinear, nonconcave function of the transmit power and channel conditions.

In any timeslot, the total network flow traversing a link l cannot exceed its link capacity; that is, the channel capacity constraint $\sum_{s:l\in L(s)} f_s \le C_l(p_l)$ exists. Because the node power is limited, $\sum_{l\in\mathcal{O}(n)} p_l \le p_n^{\max}$ must hold, where $p_n^{\max}$ denotes the maximal nodal power, and $\mathcal{O}(n)$ denotes the output link set of node n.

C. OPTIMIZATION PROBLEM

As in the general NUM problem, we use the utility $U(f_s) = \log(f_s)$ [14], which is proved twice continuously differentiable, nondecreasing, and strictly concave, to achieve congestion control and proportional fairness among the flows. By associating the objective and the constraints together, we have the following formulations:

$$\text{Maximize} \quad \sum_{s=1}^{S} U(f_s)$$

$$\text{Subject to} \quad \sum_{s:l\in L(s)} f_s \le C_l(p_l), \quad \forall l, \tag{1}$$

$$\sum_{l\in\mathcal{O}(n)} p_l \le p_n^{\max}, \quad \forall n, \tag{2}$$

$$f_s \geqslant 0, \quad \forall s, p_l \geqslant 0, \ \forall l, \tag{3}$$

$$C_l(p_l) = B\log\left(1 + \frac{p_l|h_l|^2}{d_l^\alpha}\right). \tag{4}$$

Note that the above optimization problem contains real variables F and a nonlinear, nonconcave function of variables P, so it is a complex, nonlinear programming problem. The constraint (4) induces the nonconvex property, and the log-transformation [9] is adopted to transform the constraint into a linear one so that the centralized algorithm can be applied to achieve the global optimality of the proposed algorithm. However, in the centralized method, each node needs to notify the central node of all its state information, and the central node then needs to send the allocated results to all other nodes. This will lead to an immense communication overhead, which is expensive in terms of time and resources, so it is impractical, especially for a large network. To solve this problem, a completely distributed second-order algorithm is described in detail in the next section.

III. JOINT CONGESTION CONTROL AND POWER ALLOCATION ALGORITHM

Define $\mathbf{y} = [f_1, \ldots, f_S, p_1, \ldots, p_L]^\top$ to group all the flow rate and link power variables, $\mathbf{M} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} \end{bmatrix}$ to group all network topology information, and $\mathbf{e} = [C_1, \ldots, C_L, p_1^{\max}, \ldots, p_N^{\max}]^\top \in \mathbb{R}^{(L+N-D)\times 1}$ to group all the link-capacity and node-power (removing the nodes without output link) variables; $\mathbf{0}$ represents an all-zero matrix, whose dimension is determined based on the context; $\top$ denotes the transpose of matrix. Therefore, constraints (1) and (2) can be compactly written as:

$$\mathbf{My} \le \mathbf{e}. \tag{5}$$

In order to solve the original problem, we first use the barrier method to transform the optimization problem into an unconstrained one. The reformulated problem can be written as:

$$\text{Minimize} \quad \hat{f}_\mu(\mathbf{y}). \tag{6}$$

The barrier objective function $\hat{f}_\mu(\mathbf{y})$ is as follows:

$$\begin{aligned} \hat{f}_\mu(\mathbf{y}) &= \left\{ \begin{array}{c} -\mu \sum_{s=1}^{S} U(f_s) - \sum_{s=1}^{S} \log(f_s) - \sum_{s=1}^{S} \log(p_l) \\ -\sum_{l=1}^{L} \log\left(C_l(p_l) - \sum_{s:l\in L(s)} f_s\right) \\ -\sum_{n\neq \text{Dst}(s),\forall s} \log\left(p_n^{\max} - \sum_{l\in\mathcal{O}(n)} p_l\right) \end{array} \right\} \\ &= \left\{ \begin{array}{c} \hat{f}_\mu(\mathbf{y}) - \sum_{l=1}^{L} \log\left(C_l(p_l) - \sum_{s:l\in L(s)} f_s\right) \\ -\sum_{n\neq \text{Dst}(s),\forall s} \log\left(p_n^{\max} - \sum_{l\in\mathcal{O}(n)} p_l\right) \end{array} \right\}, \end{aligned} \tag{7}$$

where

$$\hat{f}_\mu(\mathbf{y}) = -\mu \sum_{s=1}^{S} U(f_s) - \sum_{s=1}^{S} \log(f_s) - \sum_{l=1}^{L} \log(p_l), \tag{8}$$

and the barrier parameter $\mu > 0$ is used to adjust the degree of approximation with the original optimization problem.

The larger μ is, the closer the solution is to that of the original problem.

Next, take the first derivative of $\hat{f}_\mu(\mathbf{y})$ with respect to f_s and p_l respectively, and then set the results equal to zero. We thus obtain

$$\frac{\partial \hat{f}_\mu(\mathbf{y})}{\partial f_s} = -\mu U'(f_s) - \frac{1}{f_s} + \sum_{l=1}^{L} \frac{1_s(l)}{\left(C_l(p_l) - \sum\limits_{s:l\in L(s)} f_s\right)} = 0, \tag{9}$$

$$\frac{\partial \hat{f}_\mu(\mathbf{y})}{\partial p_l} = -\frac{C_l{}'(p_l)}{\left(C_l(p_l) - \sum\limits_{s:l\in L(s)} f_s\right)} - \frac{1}{p_l} + \frac{1}{\left(p_{\mathrm{Tx}(l)}^{\max} - \sum\limits_{j\in\mathcal{O}(\mathrm{Tx}(l))} p_j\right)} = 0, \tag{10}$$

where $(1_s)_l = \begin{cases} 1, & \text{if } l \in L(s), \\ 0, & \text{otherwise.} \end{cases}$

According to the primal-dual interior-point method, we define dual variables $w_l = \left(C_l(p_l) - \sum\limits_{s:l\in L(s)} f_s\right)^{-1}$ and $\varphi_n = \left(p_n^{\max} - \sum\limits_{l\in\mathcal{O}(n)} p_l\right)^{-1}$ based on (9) and (10), respectively. $\left(C_l(p_l) - \sum\limits_{s:l\in L(s)} f_s\right)$ denotes the remaining capacity on link l, which reflects the use-cost of link l, and $\left(p_n^{\max} - \sum\limits_{l\in\mathcal{O}(n)} p_l\right)$ indicates the remaining power of node n, which implies the power cost of node n. Thus, w_l and φ_n are regarded as link-congestion price and node-power price, respectively. Without loss of generality, $\mathbf{w} = [w_l, \forall l]^\top \in \mathbb{R}^{L\times 1}$, $\boldsymbol{\varphi} = [\varphi_n, \forall n]^\top \in \mathbb{R}^{(N-D)\times 1}$ and $\lambda = [\mathbf{w}^\top, \boldsymbol{\varphi}^\top]^\top$ are defined to group all link-congestion prices, node-power prices and dual variables, respectively. Therefore, we obtain the following perturbed KKT system of reformulated optimization problem, which contains stationarity (ST), primal feasibility (PF), dual feasibility (DF) and perturbed complementary slackness (CS) conditions:

$$\text{(ST): } \nabla f_\mu(\mathbf{y}) + (\mathbf{M}^\top - \nabla\mathbf{e})\boldsymbol{\lambda} = \mathbf{0}, \tag{11}$$

$$\text{(PF): } \mathbf{y} > \mathbf{0}, \mathbf{My} - \mathbf{e} < \mathbf{0}, \tag{12}$$

$$\text{(DF): } \boldsymbol{\lambda} > \mathbf{0}, \tag{13}$$

$$\text{(CS): } -\mathrm{Diag}\{\mathbf{My} - \mathbf{e}\}\boldsymbol{\lambda} = \mathbf{1}, \tag{14}$$

where $\mathbf{1}$ represents an all-one matrix, whose dimension is determined from the context.

By using Newton's method to solve the nonlinear system composed of the perturbed KKT conditions, we have

$$\begin{bmatrix} \mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\mathbf{M}^\top - \nabla\mathbf{e}_{[t]} \\ -\boldsymbol{\Lambda}_{[t]}\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) - \mathbf{Q}_{[t]} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{y}_{[t]} \\ \Delta\boldsymbol{\lambda}_{[t]} \end{bmatrix} = -\begin{bmatrix} \mathbf{g}_{[t]} + \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)\lambda_{[t]} \\ -\left(\boldsymbol{\lambda}_{[t]}\mathbf{Q}_{[t]} + \mathbf{I}\right)\mathbf{1} \end{bmatrix}, \tag{15}$$

where $\mathrm{g}_{[t]} = \nabla f_\mu(\mathbf{y}_{[t]})$ and $\mathrm{H}_{[t]} = \nabla^2 f_\mu(\mathbf{y}_{[t]})$ denote the gradient vector and Hessian matrix of $f_\mu(\mathbf{y}_{[t]})$, respectively, $\nabla\mathbf{e}_{[t]} = \frac{\partial\mathbf{e}_{[t]}}{\partial\mathbf{y}_{[t]}}$, $\nabla^2\mathbf{e}_{[t]} = \frac{\partial(\nabla\mathbf{e}_{[t]})\boldsymbol{\lambda}_{[t]}}{\partial\mathbf{y}_{[t]}}$, $\boldsymbol{\lambda}_{[t]} = \mathrm{Diag}\{\boldsymbol{\lambda}_{[t]}\}$, and $\mathrm{Q}_{[t]} = \mathrm{Diag}\{\mathbf{My}_{[t]} - \mathbf{e}_{[t]}\}$, $\mathrm{Diag}\{*\}$ indicate diagonalization, and I represents the unit matrix, whose dimension is determined from the context. We define this system a modified KKT system, where $\nabla\mathbf{e}_{[t]}$, $\nabla^2\mathbf{e}_{[t]}$ are modified matrices decided by the dynamic channel capacity constraint, so the system is adaptive to different channel capacity models. When the static channel capacity model (i.e. channel capacity is a constant) is adopted, the modified matrices $\nabla\mathbf{e}_{[t]}$ and $\nabla^2\mathbf{e}_{[t]}$ become full zero matrices, and the modified KKT system is reduced to the general KKT system (such as the KKT system in [11]). The modified KKT system is more inclusive.

According to (15), the primal Newton direction $\Delta\mathbf{y}_{[t]}$ and the dual Newton direction $\Delta\boldsymbol{\lambda}_{[t]}$ can be calculated easily:

$$\Delta\mathbf{y}_{[t]} = -\mathbf{F}_{[t]}^{-1}\left[\mathbf{g}_{[t]} - \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)\mathbf{Q}_{[t]}^{-1}\mathbf{1}\right], \tag{16}$$

$$\Delta\boldsymbol{\lambda}_{[t]} = \left\{-\mathbf{G}_{[t]}^{-1}\mathbf{1}\begin{bmatrix} \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \\ \times\left(\mathbf{g}_{[t]} + \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}_{[t]}\right) \\ -\left(\mathbf{Q}_{[t]} + \boldsymbol{\Lambda}_{[t]}^{-1}\right)\mathbf{1} \end{bmatrix}\right\}, \tag{17}$$

where

$$\mathbf{F}_{[t]} = \begin{bmatrix} \mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]} - \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right) \\ \times\mathbf{Q}_{[t]}^{-1}\boldsymbol{\Lambda}_{[t]}\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) \end{bmatrix}, \tag{18}$$

$$\mathbf{G}_{[t]} = \begin{bmatrix} \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}\right)^{-1} \\ \times\left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right) - \boldsymbol{\Lambda}_{[t]}^{-1}\mathbf{Q}_{[t]} \end{bmatrix}. \tag{19}$$

Then, with $\Delta\mathbf{y}_{[t]}$ and $\Delta\boldsymbol{\lambda}_{[t]}$, the primal variable $\mathbf{y}$ and dual variable $\boldsymbol{\lambda}$ can be iteratively updated according to the following equations:

$$\begin{bmatrix} \mathbf{y}_{[t+1]} \\ \boldsymbol{\lambda}_{[t+1]} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_{[t]} \\ \boldsymbol{\lambda}_{[t]} \end{bmatrix} + \pi_{[t]}\begin{bmatrix} \Delta\mathbf{y}_{[t]} \\ \Delta\boldsymbol{\lambda}_{[t]} \end{bmatrix}, \tag{20}$$

where $\pi_{[t]}$ is a step size.

Thus far, we have obtained the update method for the primal variable $\mathbf{y}$ and the dual variable $\boldsymbol{\lambda}$. But note that the inverse operations of matrices $\mathbf{F}_{[t]}$ and $\mathbf{G}_{[t]}$ in (16) and (17), respectively, need to collect global information such as flow rate, link power, and channel state information for the entire network; thus, only a centralized computation can be applied here. Unfortunately, the signaling overhead grows with the network size in a centralized computation pattern. Therefore, we need to further decouple these parameters so that we can update them distributedly.

A. DISTRIBUTED SOLUTION PROCEDURE

We first consider dual-variable updating with a full Newton step, that is, $\tilde{\boldsymbol{\lambda}}_{[t+1]} = \boldsymbol{\lambda}_{[t]} + \Delta\boldsymbol{\lambda}_{[t]}$. By substituting (17) into the aforementioned equation and simplifying it, we have

$$\tilde{\boldsymbol{\lambda}}_{[t+1]} = \mathbf{G}_{[t]}^{-1}\Big[-\Big(\mathbf{M}-\nabla\mathbf{e}_{[t]}^{\top}\Big) \times \Big(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\Big)^{-1}\mathbf{g}_{[t]}+\boldsymbol{\Lambda}_{[t]}^{-1}\mathbf{1}\Big]. \quad (21)$$

According to (15), we can then obtain

$$\Big(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\Big)\Delta\mathbf{y}_{[t]}+\Big(\mathbf{M}^{\top}-\nabla\mathbf{e}_{[t]}\Big)\Delta\boldsymbol{\lambda}_{[t]} = -\Big[\mathbf{g}_{[t]}+\Big(\mathbf{M}^{\top}-\nabla\mathbf{e}_{[t]}\Big)\boldsymbol{\lambda}_{[t]}\Big].$$

By combining the above equation with (21), the primal and dual Newton directions in (16) and (17) can be alternatively computed as follows:

$$\Delta\mathbf{y}_{[t]} = -\Big(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\Big)^{-1} \times\Big[\mathbf{g}_{[t]}+\Big(\mathbf{M}^{\top}-\nabla\mathbf{e}_{[t]}\Big)\tilde{\boldsymbol{\lambda}}_{[t+1]}\Big], \quad (22)$$

$$\Delta\boldsymbol{\lambda}_{[t]} = \tilde{\boldsymbol{\lambda}}_{[t+1]}-\boldsymbol{\lambda}_{[t]}. \quad (23)$$

1) DISTRIBUTED COMPUTATION OF THE PRIMAL NEWTON DIRECTION

In order to decouple the variables in (22), $\big(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\big)^{-1} = \big(\nabla^2 f_{\mu}(\mathbf{y}_{[t]})-\nabla^2\mathbf{e}_{[t]}\big)^{-1}$ first needs to be solved in a distributed manner. According to (8) and (4), we have the following (for simplicity, we omit time-slot indexes "t" in the next part):

$$\frac{\partial f_{\mu}(\mathbf{y})}{\partial f_s} = -\mu U'(f_s)-\frac{1}{f_s},$$
$$\frac{\partial^2 f_{\mu}(\mathbf{y})}{\partial f_s^2} = -\mu U''(f_s)+\frac{1}{f_s^2},$$
$$\frac{\partial f_{\mu}(\mathbf{y})}{\partial p_l} = -\frac{1}{p_l},$$
$$\frac{\partial^2 f_{\mu}(\mathbf{y})}{\partial p_l^2} = \frac{1}{p_l^2},$$

where $U'(f_s)$ and $U''(f_s)$ are the first and second derivatives of the utility function, respectively, and

$$\nabla\mathbf{e} = \frac{\partial\mathbf{e}}{\partial\mathbf{y}} = \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{C}' & \mathbf{0}\end{bmatrix},\quad \nabla\mathbf{e}\in\mathbb{R}^{(S+L)\times(L+N-D)},$$
$$\nabla^2\mathbf{e} = \frac{\partial(\nabla\mathbf{e})\lambda}{\partial\mathbf{y}} = \begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}''\boldsymbol{\Lambda}\end{bmatrix},\quad \nabla^2\mathbf{e}\in\mathbb{R}^{(S+L)\times(S+L)},$$
$$\mathbf{C}' = \mathrm{Diag}\left\{C_l' = \frac{\partial C_l}{\partial p_l},\forall l\right\},$$
$$\mathbf{C}'' = \mathrm{Diag}\left\{C_l'' = \frac{\partial^2 C_l}{\partial p_l^2},\forall l\right\},$$
$$\mathbf{C}''\boldsymbol{\Lambda} = \mathrm{Diag}\left\{w_l C_l'',\forall l\right\},$$

where C'_l and C''_l represent the first and second derivatives of the channel capacity function with respect to power p_l, respectively. Therefore, $\mathbf{g}_{[t]} = [-\mu U'(f_{1,[t]})-\frac{1}{f_{1,[t]}},\ldots,-\mu U'(f_{S,[t]})-\frac{1}{f_{S,[t]}},-\frac{1}{p_{1,[t]}},\ldots,-\frac{1}{p_{L,[t]}}]^{\top}\in\mathbb{R}^{(S+L)\times 1}$.

Furthermore, $(\mathrm{H}_{[t]}-\nabla^2\mathbf{e}_{[t]})$ can be expressed as the following diagonal structure: $(\mathrm{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}) = \begin{bmatrix}\mathbf{S}_{[t]} & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{[t]}\end{bmatrix}$, where $\mathbf{S}_{[t]} = \mathrm{Diag}\{-\mu U''(f_{s,[t]})+\frac{1}{f_{s,[t]}^2},\forall s\}$, $\mathbf{P}_{[t]} = \mathrm{Diag}\{\frac{1}{p_{l,[t]}^2}-w_{l,[t]}C''_{l,[t]},\forall l\}$. $\mathbf{S}_{[t]}$ and $\mathbf{P}_{[t]}$ are both diagonal matrices, so it is easy to compute their own inverse matrices:

$$\mathbf{S}_{[t]}^{-1} = \mathrm{Diag}\left\{\left(-\mu U''\left(f_{s,[t]}\right)+\frac{1}{f_{s,[t]}^2}\right)^{-1},\forall s\right\},$$
$$\mathbf{P}_{[t]}^{-1} = \mathrm{Diag}\left\{\left(\frac{1}{p_{l,[t]}^2}-w_{l,[t]}C''_{l,[t]}\right)^{-1},\forall l\right\}.$$

Finally, the inverse matrix of $(\mathrm{H}_{[t]}-\nabla^2\mathbf{e}_{[t]})$ can be written as

$$\Big(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\Big)^{-1} = \begin{bmatrix}\mathbf{S}_{[t]}^{-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{P}_{[t]}^{-1}\end{bmatrix}.$$

By substituting the above result into (22), we can obtain distributed update formulas for the Newton directions of the flow rate and link power, respectively:

$$\Delta f_{s,[t]} = -\left(-\mu U''\left(f_{s,[t]}\right)+\frac{1}{f_{s,[t]}^2}\right)^{-1} \times\left[-\mu U'\left(f_{s,[t]}\right)-\frac{1}{f_{s,[t]}}+\sum_{l\in L(s)}\tilde{w}_{l,[t+1]}\right], \quad (24)$$

$$\Delta p_{l,[t]} = -\left(\frac{1}{p_{l,[t]}^2}-w_{l,[t]}C''_{l,[t]}\right)^{-1} \times\left[-\frac{1}{p_l}-C'_{l,[t]}\tilde{w}_{l,[t+1]}+\tilde{\varphi}_{\mathrm{Tx}(l),[t+1]}\right]. \quad (25)$$

2) DISTRIBUTED COMPUTATION OF THE DUAL NEWTON DIRECTION

The limitation of distributed computing of the dual variables in (21) is that $\mathbf{G}_{[t]}^{-1}$ requires global information. To solve this problem, we first transform (21) into the following linear equation, which can be solved with the matrix-splitting method [15]:

$$\mathbf{G}_{[t+1]}\tilde{\boldsymbol{\lambda}}_{[t+1]} = -\Big(\mathbf{M}-\nabla\mathbf{e}_{[t]}^{\top}\Big) \times\Big(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\Big)^{-1}\mathbf{g}_{[t]}+\boldsymbol{\Lambda}_{[t]}^{-1}\mathbf{1}. \quad (26)$$

Matrix splitting is a common way to solve linear equations in an iterative method, so we use it to decompose $\mathbf{G}_{[t]}$. As a result, we only need to invert a diagonal matrix in the calculation process of $\tilde{\boldsymbol{\lambda}}_{[t+1]}$.

The matrix-splitting method can be simply summarized as follows. Consider a consistent linear equation system $\mathbf{Fz} = \mathbf{d}$, where $\mathbf{F}\in\mathbb{R}^{n\times n}$ is a nonsingular matrix and $\mathbf{z},\mathbf{d}\in\mathbb{R}^{n}$. Suppose that $\mathbf{F}$ is split into a nonsingular matrix $\mathbf{F}_1$ and another matrix $\mathbf{F}_2$ according to $\mathbf{F} = \mathbf{F}_1 - \mathbf{F}_2$. Let $\mathbf{z}^0$ be an arbitrary starting vector. Then, a sequence of approximate solutions can be generated using the following iterative scheme:

$$\mathbf{z}^{k+1}=(\mathbf{F}_1^{-1}\mathbf{F}_2)\mathbf{z}^{k}+\mathbf{F}_1^{-1}\mathbf{d},\ k \geq 0. \tag{27}$$

Generally, $\mathbf{F}_1$ should be an easily invertible matrix (e.g., a diagonal matrix). It can be shown that this iterative method is convergent to the unique solution $\mathbf{z} = \mathbf{F}^{-1}\mathbf{d}$ with $k \to \infty$ if and only if the spectral radius $\rho(\mathbf{F}_1^{-1}\mathbf{F}_2)$ of the matrix $\mathbf{F}_1^{-1}\mathbf{F}_2$ satisfies $\rho(\mathbf{F}_1^{-1}\mathbf{F}_2) < 1$.

Based on the above description, we can obtain the iterative calculation formula of $\tilde{\boldsymbol{\lambda}}_{[t+1]}$. The specific details are given in Theorem 1 below. After Theorem 1 is given, we prove that it satisfies the convergence condition that the spectral radius $\rho(\mathbf{F}_1^{-1}\mathbf{F}_2) < 1$ in the splitting process.

Theorem 1: Split $\mathbf{G}_{[t]}$ as

$$\mathbf{G}_{[t]} = \left(\boldsymbol{\Phi}_{[t]} + \beta\bar{\boldsymbol{\Omega}}_{[t]}\right) - \left(\beta\bar{\boldsymbol{\Omega}}_{[t]} - \boldsymbol{\Omega}_{[t]}\right),$$

where $\boldsymbol{\Phi}_{[t]} = \mathrm{Diag}\{\mathbf{G}_{[t]}\}$ is a diagonal matrix having the same main diagonal of $\mathbf{G}_{[t]}$, $\boldsymbol{\Omega}_{[t]} = \mathbf{G}_{[t]} - \boldsymbol{\Phi}_{[t]}$ denotes the matrix containing entries after subtracting $\boldsymbol{\Phi}_{[t]}$ from $\mathbf{G}_{[t]}$, $\bar{\boldsymbol{\Omega}}_{[t]}$ is a diagonal matrix with the diagonal entries defined by $(\bar{\boldsymbol{\Omega}}_{[t]})_{ii} = \sum_j \left|(\boldsymbol{\Omega}_{[t]})_{ij}\right|$, and $\beta > \frac{1}{2}$ is a parameter used to coordinate convergence performance. Then, $\tilde{\boldsymbol{\lambda}}_{[t+1]}$ can be solved by the following iterative formula:

$$\begin{aligned}\tilde{\boldsymbol{\lambda}}_{[t+1]}^{k+1} &= \left(\boldsymbol{\Phi}_{[t]} + \beta\bar{\boldsymbol{\Omega}}_{[t]}\right)^{-1}\left(\beta\bar{\boldsymbol{\Omega}}_{[t]} - \boldsymbol{\Omega}_{[t]}\right)\tilde{\boldsymbol{\lambda}}_{[t]}^{k} \\ &\quad + \left(\boldsymbol{\Phi}_{[t]} + \beta\bar{\boldsymbol{\Omega}}_{[t]}\right)^{-1} \\ &\quad \times \left[-\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\mathbf{g}_{[t]} + \boldsymbol{\Lambda}_{[t]}^{-1}\mathbf{1}\right].\end{aligned} \tag{28}$$

And $\tilde{\boldsymbol{\lambda}}_{[t+1]}^{k+1}$ converges to the solution

$$\tilde{\boldsymbol{\lambda}}_{[t+1]} = \mathbf{G}_{[t]}^{-1}\left[-\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)\mathbf{g}_{[t]} + \boldsymbol{\Lambda}_{[t]}^{-1}\mathbf{1}\right]$$

as $k \to \infty$.

To prove Theorem 1 holds, we need to prove that $\rho((\boldsymbol{\Phi}_{[t]} + \beta\bar{\boldsymbol{\Omega}}_{[t]})^{-1}(\beta\bar{\boldsymbol{\Omega}}_{[t]} - \boldsymbol{\Omega}_{[t]})) < 1$ holds. To do so, we first introduce the two following lemmas, which will be used in the proof:

Lemma 1: Suppose that $\mathbf{F}$ is a real symmetric matrix. If both matrices $\mathbf{F}_1 - \mathbf{F}_2$ and $\mathbf{F}_1 + \mathbf{F}_2$ are positive definite, then $\rho(\mathbf{F}_1^{-1}\mathbf{F}_2) < 1$.

Lemma 2: If a symmetric matrix Q is strictly diagonally dominant, that is, $|Q_{ii}| > \sum_{i \neq j}|Q_{ij}|$, and if $Q_{ii} > 0$ for all i, then Q is positive definite.

Convergence proof for Theorem 1:

Step (1): To prove that $\mathbf{G}_{[t]}$ is a real symmetric matrix. Expanding $\mathbf{G}_{[t]}$ according to (19), we have

$$\begin{aligned}\mathbf{G}_{[t]} &= \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \\ &\quad \times \left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right) - \boldsymbol{\lambda}_{[t]}^{-1}Q_{[t]} \\ &= \begin{bmatrix}\mathbf{R} & -\mathbf{C}'_{[t]} \\ \mathbf{0} & \mathbf{T}\end{bmatrix}\begin{bmatrix}\mathbf{S}_{[t]}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{P}_{[t]}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{R}^{\top} & \mathbf{0} \\ -\left(\mathbf{C}'_{[t]}\right)^{\top} & \mathbf{T}^{\top}\end{bmatrix} \\ &\quad - \mathrm{Diag}\left\{\boldsymbol{\lambda}_{[t]}^{-1}\right\} \times \mathrm{Diag}\left\{\mathbf{M}\mathbf{y}_{[t]} - \mathbf{e}_{[t]}\right\} \\ &= \begin{bmatrix}\mathbf{R}\mathbf{S}_{[t]}^{-1}\mathbf{R}^{\top} + \mathbf{C}'_{[t]}\mathbf{P}^{-1}\left(\mathbf{C}'_{[t]}\right)^{\top} & -\mathbf{C}'_{[t]}\mathbf{P}_{[t]}^{-1}\mathbf{T}^{\top} \\ -\mathbf{T}\mathbf{P}_{[t]}^{-1}\left(\mathbf{C}'_{[t]}\right)^{\top} & \mathbf{T}\mathbf{P}_{[t]}^{-1}\mathbf{T}^{\top}\end{bmatrix} \\ &\quad + \mathrm{Diag}\left\{\left(\boldsymbol{\lambda}_{[t]}^{-1}\right)^{\top}\left(\mathbf{e}_{[t]} - \mathbf{M}\mathbf{y}_{[t]}\right)\right\}.\end{aligned}$$

The second item in the above formula:

$$\left(\mathrm{Diag}\left\{\left(\boldsymbol{\lambda}_{[t]}^{-1}\right)^{\top}\left(\mathbf{e}_{[t]} - \mathbf{M}\mathbf{y}_{[t]}\right)\right\}\right)_{ii} = \begin{cases} w_i^{-1}\left(\mathrm{C}_{\mathrm{rest}}\right)_i, & i \leq L \\ \varphi_i^{-1}\left(\mathrm{P}_{\mathrm{rest}}\right)_i, & L+1 \leq i \leq L+N-D\end{cases}$$

where the matrices $\mathrm{C}_{\mathrm{rest}} \in \mathbb{R}^{L\times 1}$ and $\mathrm{P}_{\mathrm{rest}} \in \mathbb{R}^{(N-D)\times 1}$ are the 1st to Lth item and the $(L+1)$ th item to the $(L+N-D)$ th item of $(\mathbf{e}_{[t]} - \mathbf{M}\mathbf{y}_{[t]})$, respectively, $\mathrm{C}_{\mathrm{rest}}$ denotes the remaining capacity of each link, and $\mathrm{P}_{\mathrm{rest}}$ denotes the remaining power of each node. Obviously, $\mathbf{G}_{[t]}$ is a real symmetric matrix.

Step (2): To prove that $\mathbf{G}_{[t]}$ is positive definite.

Because the diagonal elements of the diagonal matrix $\mathrm{H}_{[t]}$ are positive, so it is a positive definite matrix. Similarly, the diagonal elements of the diagonal matrix $-\nabla^2\mathbf{e}_{[t]}$ are nonnegative (because the second-order derivative of the channel capacity function (4) with respect to power is negative), so $\mathrm{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}$ is strictly diagonally dominant. From Lemma 2, we can derive that $\mathrm{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}$ is a positive definite. According to the nature of the positive definite matrix, $(\mathrm{H}_{[t]} - \nabla^2\mathbf{e}_{[t]})^{-1}$ is also positive definite, and all the eigenvalues are greater than zero. Regarding $(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top})(\mathrm{H}_{[t]} - \nabla^2\mathbf{e}_{[t]})^{-1}(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]})$, because it is an invertible matrix, we have

$$\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right)\mathbf{z} \neq \mathbf{0}.$$

Therefore, we can find that

$$\begin{aligned}&\mathbf{z}^{\top}\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right)\mathbf{z} \\ &= \left[\left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right)\mathbf{z}\right]^{\top}\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\left[\left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right)\mathbf{z}\right] \\ &\geq \frac{1}{\lambda_{\min}\left\{\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)\right\}}\left\|\left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right)\mathbf{z}\right\|^2 \\ &> 0.\end{aligned}$$

Hence, $\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right)$ is positive definite. Simultaneously, the diagonal matrix $\left(-\boldsymbol{\lambda}_{[t]}^{-1}Q_{[t]}\right)$ is strictly diagonally dominant. From Lemma 2, it is a positive definite matrix. As a conclusion,

$$\begin{aligned}\mathbf{G}_{[t]} &= \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \\ &\quad \times \left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right) - \boldsymbol{\Lambda}_{[t]}^{-1}Q_{[t]}\end{aligned}$$

is also positive definite.

Step (3): To prove that

$$\rho\left(\left(\mathbf{\Phi}_{[t]}+\beta\bar{\mathbf{\Omega}}_{[t]}\right)^{-1}\left(\beta\bar{\mathbf{\Omega}}_{[t]}-\mathbf{\Omega}_{[t]}\right)\right)<1$$

holds. First, $\mathbf{G}_{[t]}$ is positive definite. Hence,

$$(\mathbf{\Phi}_{[t]}+\beta\bar{\mathbf{\Omega}}_{[t]})-(\beta\bar{\mathbf{\Omega}}_{[t]}-\mathbf{\Omega}_{[t]})=\mathbf{\Phi}_{[t]}+\mathbf{\Omega}_{[t]}$$

is positive definite. Next, we take

$$(\mathbf{\Phi}_{[t]}+\beta\bar{\mathbf{\Omega}}_{[t]})+(\beta\bar{\mathbf{\Omega}}_{[t]}-\mathbf{\Omega}_{[t]})=\mathbf{\Phi}_{[t]}+2\beta\bar{\mathbf{\Omega}}_{[t]}-\mathbf{\Omega}_{[t]}$$

into consideration. All diagonal entries in $\mathbf{\Phi}_{[t]}$ are positive, so $\mathbf{\Phi}_{[t]}$ is positive definite. On the other hand, the definitions of $\bar{\mathbf{\Omega}}_{[t]}$ and $\mathbf{\Omega}_{[t]}$ indicate that the entries of each row in $2\beta\bar{\mathbf{\Omega}}_{[t]}-\mathbf{\Omega}_{[t]}$ satisfy

$$\begin{aligned}(2\beta\bar{\mathbf{\Omega}}_{[t]}-\mathbf{\Omega}_{[t]})_{ii}&-\sum_{j\neq i}\left|(2\beta\bar{\mathbf{\Omega}}_{[t]}-\mathbf{\Omega}_{[t]})_{ij}\right|\\&=(2\beta-1)\sum_{j\neq i}\left|(\mathbf{\Omega}_{[t]})_{ij}\right|>0,\end{aligned}$$

for $\beta>\frac{1}{2}$. Similarly, it is clear that $2\beta\bar{\mathbf{\Omega}}_{[t]}-\mathbf{\Omega}_{[t]}$ is strictly diagonally dominant and hence positive definite. In conclusion, $\mathbf{\Phi}_{[t]}+2\beta\bar{\mathbf{\Omega}}_{[t]}-\mathbf{\Omega}_{[t]}$ is also positive definite. Because $\mathbf{\Phi}_{[t]}+\beta\bar{\mathbf{\Omega}}_{[t]}$ and $\beta\bar{\mathbf{\Omega}}_{[t]}-\mathbf{\Omega}_{[t]})$ are both positive definite, according to Lemma 1, $\rho((\mathbf{\Phi}_{[t]}+\beta\bar{\mathbf{\Omega}}_{[t]})^{-1}(\beta\bar{\mathbf{\Omega}}_{[t]}-\mathbf{\Omega}_{[t]}))<1$ holds, and the proof of Theorem 1 is completed. Associating Theorem 1 with the expanded structure of matrix $\mathbf{G}_{[t]}$, we can obtain the distributed calculation formula of link-congestion price $\tilde{w}$ and node-power price $\tilde{\varphi}$ at a time slot as follows (for convenience, the time-slot indexes are omitted; for each equation, the variables on the left of the equals sign represent the values at the $(t+1)$ th time-slot and the variables on the right are those at the t th time-slot):

$$\begin{aligned}\tilde{w}_l^{k+1}&\\=&\Bigg\{\sum_s\mathbf{R}_{ls}\mathbf{S}_s^{-1}+(C_l')^2\mathbf{P}_l^{-1}+w_l^{-1}(C_{\text{rest}})_l\\&+\beta\Bigg[\sum_s\Bigg(\mathbf{R}_{ls}\mathbf{S}_s^{-1}\sum_{i\neq l}\mathbf{R}_{is}\Bigg)-\sum_n\Big(\mathbf{T}_{nl}\mathbf{P}_l^{-1}C_l'\Big)\Bigg]\Bigg\}^{-1}\\&\times\Bigg\{\beta\Bigg[\sum_s\Bigg(\mathbf{R}_{ls}\mathbf{S}_s^{-1}\sum_{i\neq l}\mathbf{R}_{is}\Bigg)+\sum_n\Big(\mathbf{T}_{nl}\mathbf{P}_l^{-1}C_l'\Big)\Bigg]\tilde{w}_l^{k}\\&-\Bigg[\sum_{i\neq l}\sum_s\Big(\mathbf{R}_{ls}\mathbf{S}_s^{-1}\mathbf{R}_{is}\tilde{w}_i^{k}\Big)-\sum_n\Big(\mathbf{T}_{nl}\mathbf{P}_l^{-1}C_l'\tilde{\varphi}_n^{k}\Big)\Bigg]\Bigg\}\\&+\Bigg\{\sum_s\mathbf{R}_{ls}\mathbf{S}_s^{-1}+(C_l')^2\mathbf{P}_l^{-1}+w_l^{-1}(C_{\text{rest}})_l\\&+\beta\Bigg[\sum_s\Bigg(\mathbf{R}_{ls}\mathbf{S}_s^{-1}\sum_{i\neq l}\mathbf{R}_{is}\Bigg)-\sum_n\Big(\mathbf{T}_{nl}\mathbf{P}_l^{-1}C_l'\Big)\Bigg]\Bigg\}^{-1}\\&\times\Bigg[-\sum_s\Big(\mathbf{R}_{ls}\mathbf{S}_s^{-1}\mathbf{g}_l\Big)+C_l'\mathbf{P}_l^{-1}\mathbf{g}_{s+l}+w_l^{-1}\Bigg],\end{aligned}\tag{29}$$

$$\begin{aligned}\tilde{\varphi}_n^{k+1}&\\=&\Bigg\{\sum_l\mathbf{T}_{nl}\mathbf{P}_l^{-1}+\varphi_n^{-1}(P_{\text{rest}})_n\\&+\beta\Bigg[\sum_l\Bigg(\mathbf{T}_{nl}\mathbf{P}_l^{-1}\sum_{m\neq n}\mathbf{T}_{ml}\Bigg)-\sum_l\Big(\mathbf{T}_{nl}\mathbf{P}_l^{-1}C_l'\Big)\Bigg]\Bigg\}^{-1}\\&\times\Bigg\{\beta\Bigg[\sum_l\Bigg(\mathbf{T}_{nl}\mathbf{P}_l^{-1}\sum_{m\neq n}\mathbf{T}_{ml}\Bigg)-\sum_l\Big(\mathbf{T}_{nl}\mathbf{P}_l^{-1}C_l'\Big)\Bigg]\tilde{\varphi}_n^{k}\\&-\Bigg[-\sum_l\Big(\mathbf{T}_{nl}\mathbf{P}_l^{-1}C_l'\tilde{w}_l^{k}\Big)+\sum_{m\neq n}\sum_l\Big(\mathbf{T}_{nl}\mathbf{P}_l^{-1}\mathbf{T}_{ml}\tilde{\varphi}_m^{k}\Big)\Bigg]\Bigg\}\\&+\Bigg\{\sum_l\mathbf{T}_{nl}\mathbf{P}_l^{-1}+\varphi_n^{-1}(P_{\text{rest}})_n\\&+\beta\Bigg[\sum_l\Bigg(\mathbf{T}_{nl}\mathbf{P}_l^{-1}\sum_{m\neq n}\mathbf{T}_{ml}\Bigg)-\sum_l\Big(\mathbf{T}_{nl}\mathbf{P}_l^{-1}C_l'\Big)\Bigg]\Bigg\}^{-1}\\&\times\Bigg[-\sum_l\Big(\mathbf{T}_{nl}\mathbf{P}_l^{-1}\mathbf{g}_{S+l}\Big)+\varphi_n^{-1}\Bigg].\end{aligned}\tag{30}$$

As (29) and (30) indicate, all the information needed during the update process comes from the node itself and its one-hop neighbors. Finally, according to (23) and (30), the Newton directions of link-congestion price and node-power price are updated as follows:

$$\Delta w_{l,[t]}=\tilde{w}_{l,[t+1]}-w_{l,[t]},\quad\forall l,\tag{31}$$

$$\Delta\varphi_{n,[t]}=\tilde{\varphi}_{n,[t+1]}-\varphi_{n,[t]},\quad\forall n.\tag{32}$$

B. THE IMPLEMENTATION STEP OF JCCPA

It is worth noting that JCCPA does not require strict feasibility of initial values. For cases in which the iteration result exceeds the network resource limit, according to (12) and (13) in the perturbed KKT conditions, we first obtain the following feasible set of original and dual variables: $S_\varepsilon^M=\left\{(\mathbf{y},\boldsymbol{\lambda})\,\middle|\,\begin{matrix}\varepsilon\mathbf{1}\leq\mathbf{y}\leq M\mathbf{1},\boldsymbol{\lambda}\geq\varepsilon\mathbf{1}\\\mathbf{My}\leq\mathbf{e}-\varepsilon\mathbf{1}\end{matrix}\right\}$, where the constant $\varepsilon>0$ can be made arbitrarily close to zero and the constant $M>0$ is used to restrain the bursty. We then use the set projection [13] to adjust the results. The main idea of set projection is that when the iterative point exceeds the defined convex set during the iterative process, a point within the set nearest to that point is found to replace the current iteration point, that is, $\underset{\mathbf{x}\in S}{\text{Minimize}}\ \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|^2$. In the defined convex set S, a point $\mathbf{x}$, which minimizes the 2-norm value (i.e., distance) of the difference between the original iteration value and the current iteration value, is chosen to be the new iteration point. The following is a simple description of the projection process:

$$\begin{bmatrix}\mathbf{y}_{[t+1]}\\\boldsymbol{\lambda}_{[t+1]}\end{bmatrix}=\left(\begin{bmatrix}\mathbf{y}_{[t]}\\\boldsymbol{\lambda}_{[t]}\end{bmatrix}+\pi\begin{bmatrix}\Delta\mathbf{y}_{[t]}\\\Delta\boldsymbol{\lambda}_{[t]}\end{bmatrix}\right)_{S_\varepsilon^M},\tag{33}$$

where $(*)_{S_\varepsilon^M}$ represents the projection result of $(\mathbf{y},\boldsymbol{\lambda})$ on the set S_ε^M.

Thus far, the distributed implementation of JCCPA can be described as follows:

1) Initialization phase: Initialize all flow source rates $f_{s,[0]}$ and link power $p_{l,[0]}$, as well as link congestion price $w_{l,[0]}$ and node power price $\varphi_{n,[0]}$, and select the update step size $\pi \in (0, 1]$.
2) Distributed update of link-congestion price and node-power price: At time-slot t, each node updates the link-congestion price $\tilde{\mathbf{w}}_{[t+1]}$ and node-power price $\tilde{\boldsymbol{\varphi}}_{[t+1]}$ in full Newton step size according to (29) and (30).
3) Distributed update of primal Newton direction: Each node updates the Newton directions of flow rate and link power through (24) and (25).
4) Distributed update of dual Newton direction: Each node updates the Newton directions of link-congestion price and node-power price through (31) and (32).
5) Each node completes the update of the flow rate, link power, link-congestion price, and node-power price through (34)-(37):

$$f_{s,[t+1]} = f_{s,[t]} + \pi \Delta f_{s,[t]}, \tag{34}$$
$$p_{l,[t+1]} = p_{l,[t]} + \pi \Delta p_{l,[t]}, \tag{35}$$
$$w_{l,[t+1]} = w_{l,[t]} + \pi \Delta w_{l,[t]}, \tag{36}$$
$$\varphi_{n,[t+1]} = \varphi_{n,[t]} + \pi \Delta \varphi_{n,[t]}. \tag{37}$$

6) If the update result in step 5 exceeds the feasible set, use the projection result obtained through (33) as the actual allocation result.
7) If $|\mathbf{y}_{[t+1]} - \mathbf{y}_{[t]}| \leq \varepsilon \mathbf{1}$, stop the iteration; otherwise, repeat the above steps until the algorithm converges.

C. CONVERGENCE ANALYSIS

In this subsection, we discuss the convergence property of JCCPA. Consider the Lyapunov drift function $V\left(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}\right) = \frac{1}{2\pi}\left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|^2 + \frac{1}{2\mu^3\pi}\left\|\boldsymbol{\lambda}_{[t]} - \boldsymbol{\lambda}^*\right\|^2$, which denotes the distance between the primal dual variable pair($\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}$)and the perturbed KKT point($\mathbf{y}^*, \boldsymbol{\lambda}^*$). According to the update equations for the primal and dual variables and perturbed KKT conditions and combining with the derivation in [11], the Lyapunov drift rate in one time slot can be derived as

$$\begin{aligned}\Delta V\left(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}\right) &= V\left(\mathbf{y}_{[t+1]}, \boldsymbol{\lambda}_{[t+1]}\right) - V\left(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}\right) \\ &\leq -\mathrm{B}_1\left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|^2 + \pi \mathrm{B}_2 + \frac{1}{\mu}\mathrm{B}_3 + \frac{1}{\mu}\mathrm{B}_4,\end{aligned}$$

where $\mathrm{B}_1 = \frac{\lambda_{\min}\{\mathbf{H}\}}{\lambda_{\min}\{\mathbf{F}\}}$ and $\lambda_{\min}\{\mathbf{F}\} = \inf_t \left\{\lambda_{\min}\left\{\mathbf{F}_{[t]}\right\}\right\}$ are all positive numbers independent of μ. The detailed proof is provided in Appendix A.

When $\pi = O(1/\mu)$,

$$\begin{aligned}\Delta V\left(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}\right) &= V\left(\mathbf{y}_{[t+1]}, \boldsymbol{\lambda}_{[t+1]}\right) - V\left(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}\right) \\ &\leq -\mathrm{B}_1\left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|^2 + \frac{1}{\mu}\hat{\mathbf{B}},\end{aligned}$$

where $\hat{\mathbf{B}} = \upsilon \mathrm{B}_2 + \mathrm{B}_3 + \mathrm{B}_4$ and υ is a positive constant. Omitting the Lyapunov drift results from time slot 0 to $T-1$, we can obtain

$$V(\mathbf{y}_{[T]}, \boldsymbol{\lambda}_{[T]}) - V(\mathbf{y}_{[0]}, \boldsymbol{\lambda}_{[0]}) \leq -\mathrm{B}_1 \sum_{t=0}^{T-1}\left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|^2 + \frac{T}{\mu}\hat{\mathbf{B}}.$$

Dividing both sides by $T\mathrm{B}_1$ and letting $T \to \infty$, we then have

$$\frac{1}{T}\sum_{t=0}^{T-1}\left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|^2 \leq \frac{\hat{\mathrm{B}}}{\mu \mathrm{B}_1} = \frac{\tilde{\mathrm{B}}^2}{\mu},$$

where $\tilde{\mathbf{B}}^2 = \frac{\hat{\mathbf{B}}}{\mathrm{B}_1}$. Finally, according to the norm triangle inequality, taking the relationship between 1-norm and 2-norm into consideration, we can obtain

$$\left|\frac{1}{T}\sum_{t=0}^{T-1}\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)\right| \leq \left(\frac{1}{T}\sum_{t=0}^{T-1}\left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|^2\right)^{\frac{1}{2}} \leq \frac{\tilde{\mathbf{B}}}{\sqrt{\mu}}.$$

Thus far, we have shown that the primal variable will converge to within a very small neighborhood of the optimal value when $\mu \to \infty$.

IV. SIMULATION RESULTS

In this section, we present the simulation results to illustrate the performance of JCCPA. We also compare JCCPA with the first-order algorithm and other second-order algorithms.

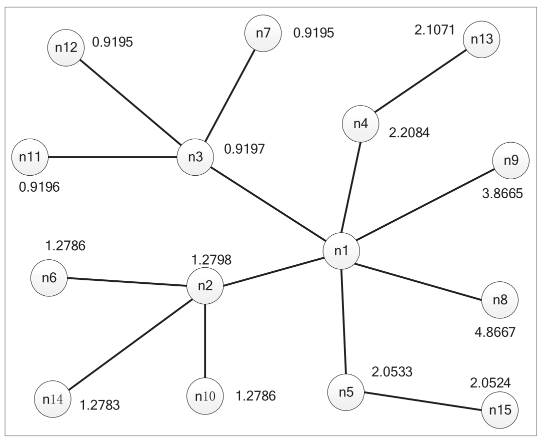


FIGURE 1. Network topology with flow rate allocation results.

In our simulation, we consider a wireless multihop network as shown in Figure 1. In this network, 15 nodes are randomly generated in a $700m \times 700m$ region. The node n1 is a gateway node, and the remaining nodes are source nodes. There are 14 flows in the network, and they converge to the gateway node and leave the network eventually. Under the same network resource configuration, the convergence rate of the proposed second-order algorithm is first compared with that of the first-order Lagrange dual decomposition algorithm [6], which optimizes the same problem and achieves almost the same utility as JCCPA. The network utility and energy efficiency of JCCPA are further compared with those of the newest second-order algorithms in [9] and [11].

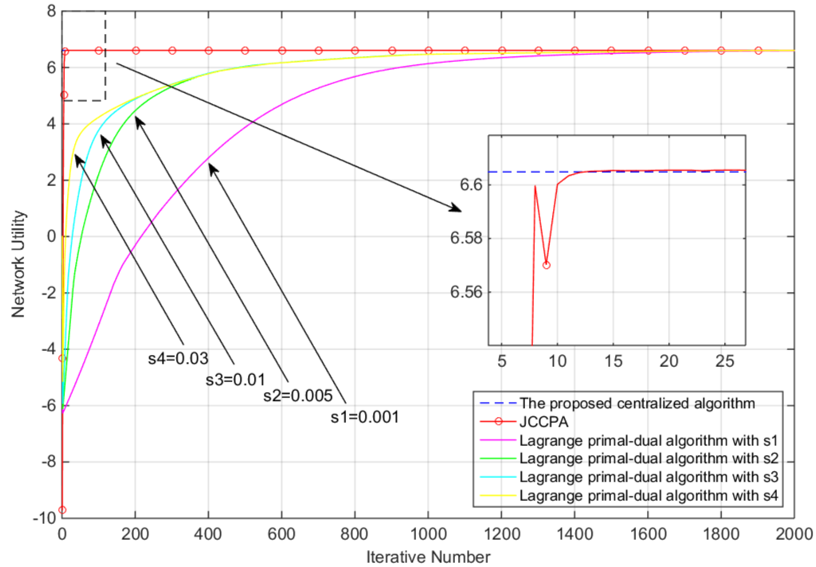


FIGURE 2. Comparison of convergence performance.

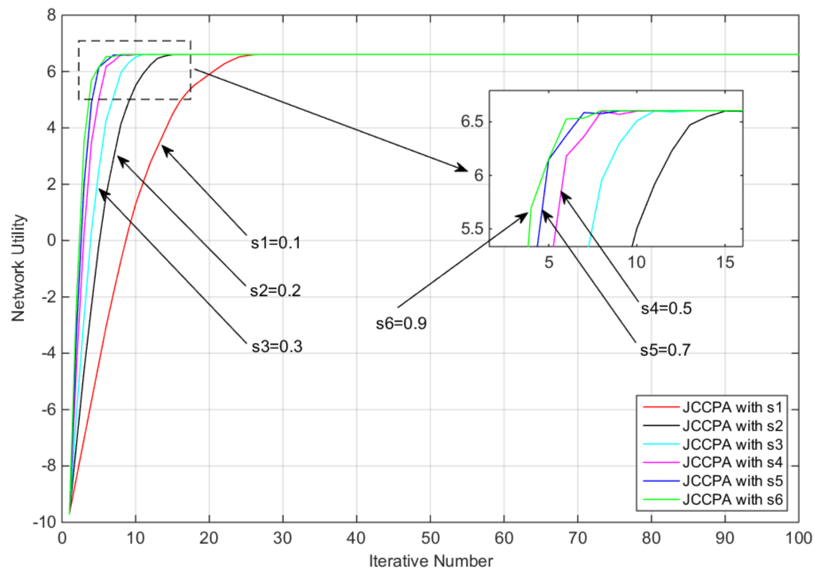


FIGURE 3. Convergence performance of JCCPA under different step sizes.

Network utility and energy utility are defined as $\sum_s \log f_s$ and $\frac{\sum_s f_s}{\sum_l p_l}$, respectively.

Simulation parameter settings are as follows. The primal and dual variables' iteration step size $\pi = 0.5$. The initialized flow rates are all 0.5Mbps, and the initialized normalized transmit power of each link is 1.8. The maximal node power is 3. The logical topology is shown in Figure 1. The allocated flow rates after optimization are labeled beside each source node.

Figure 2 compares the convergence performance among the Lagrangian dual decomposition algorithm under different iterative step sizes (s1 = 0.001, s2 = 0.005, s3 = 0.01, and s4 = 0.03), the proposed centralized and distributed algorithms. As Figure 2 shows, the convergence rate of the Lagrangian algorithm becomes increasingly fast with the increase of iterative step size, and it finally converges to the optimum at the 1500th iteration. The proposed second-order algorithm achieves the same effects at about the 18th iteration. Figure 3 further shows the convergence performance of JCCPA under different iterative step sizes (s1 = 0.1, s2 = 0.2, s3 = 0.3, s4 = 0.5, s5 = 0.7, and s6 = 0.9). The convergence speed of JCCPA becomes slower when the step size becomes smaller. However, the slowest convergence speed is guaranteed in 30th iteration. JCCPA is fast convergent as long as the step size is within the defined interval. As for the Lagrangian

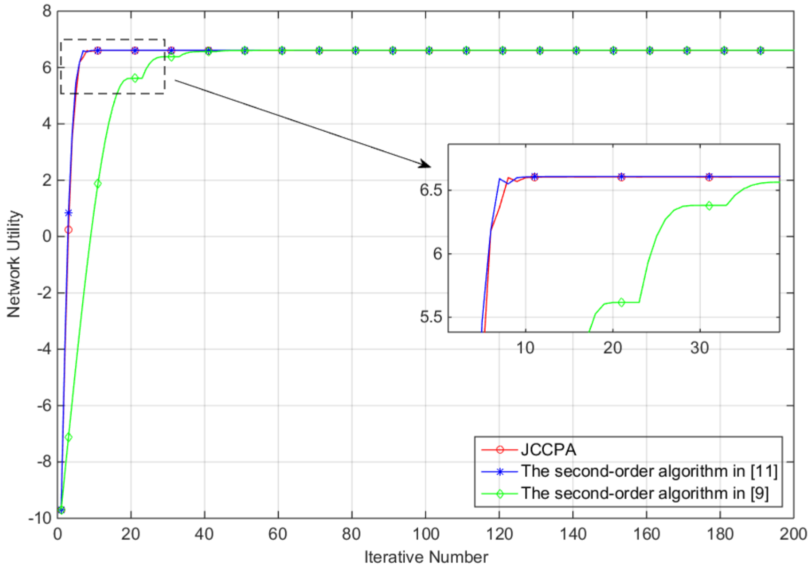


FIGURE 4. Comparison of network utility.

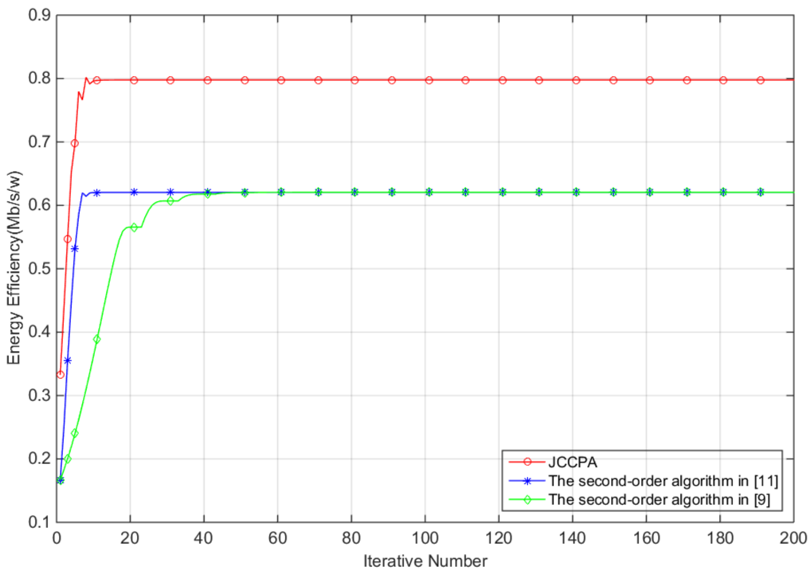


FIGURE 5. Comparison of energy efficiency.

algorithm, iteration step size is generally ambiguous, and not in a range interval, so that this selection is opportunistic and hard to find a best step size just once. It is therefore clear that although the two algorithms can converge to the same optimum, JCCPA is insensitive to the selection of step size and its convergence speed is more than 30 times faster than that of the Lagrangian algorithm.

Figure 4 compares the network utility of JCCPA with the second-order algorithm without consideration of power allocation in [9] and [11]. Figure 5 plots the energy utility curves. In the simulation, the node-power constraint for JCCPA is $p_n = 3$. JCCPA adjusts the flow rate allocation according to the congestion condition. Link flow rate further influences the transmit power allocation, and transmit power determines the link capacity and congestion condition. Notably, in the link-capacity model, the received SNR definitely depends on the transmit power on the link itself, regardless of the inter-link interference. Because the node power in [9] and [11] is always the maximum, each link capacity is keeping the maximum free in the context of interference. Under the constraint of a bottleneck link in the uplink transmission, as shown in Figure 4, the two algorithms arrive at the same network utility. However, a fixed power distribution strategy induces large power waste, and conversely, an adaptive power distribution strategy adjusts the node power on demand, reduces not only power consumption but also interlink interference, and tailors the capacity to the need. As the above discussion suggests, we can see from Figure 5 that the energy efficiency of JCCPA is 37.5% higher than the algorithm without consideration of power allocation.

V. CONCLUSION

The first-order algorithms for cross-layer resource optimization allocation in wireless multihop ad hoc networks have been the subject of a great deal of research. In order to improve the convergence rate of the first-order algorithms, a centralized second-order optimization algorithm for joint congestion control, rate allocation, and power allocation based on the primal-dual interior-point method is designed in the case that the network node power is limited and the routing is known. Based on this, matrix-splitting technology is further used to achieve the distributed implementation of the algorithm. The simulation results show that the convergence speed of the algorithm is several tens of times faster than the first-order Lagrangian algorithm. Simultaneously, compared to the newest second-order algorithm, the algorithm improves energy efficiency by 37.5%.

APPENDIX A
LYAPUNOV DRIFT RATE IN ONE TIME SLOT

In order to facilitate the derivation of one-slot Lyapunov drift rate, we first prove the following lemma.

Theorem 1: In the proposed algorithm, for a given μ, if $\|\boldsymbol{\lambda}_{[0]}\| < \infty$, then $\|\boldsymbol{\lambda}_{[t]}\| < \infty$ for all t.

Proof: Use mathematical induction to prove Theorem 1. □

For $t = 0$, $\|\boldsymbol{\lambda}_{[0]}\| < \infty$ is obviously true according to the assumption in Lemma 2. Suppose that we have $\|\boldsymbol{\lambda}_{[t]}\| < B_0 < \infty$ at time slot t. Then, we need to prove that $\|\boldsymbol{\lambda}_{[t+1]}\| < \infty$ at time slot $t+1$. First, we let $\tilde{\boldsymbol{\lambda}}_{[t+1]} = \boldsymbol{\lambda}_{[t]} + \Delta\boldsymbol{\lambda}_{[t]}$ (i.e., taking a full Newton step $\pi = 1$). According to the updating formula (21), we have

$$\tilde{\boldsymbol{\lambda}}_{[t+1]} = \Big[\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right) - \boldsymbol{\lambda}_{[t]}^{-1}\mathbf{Q}_{[t]}\Big]^{-1} \times \Big[-\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right) \times \left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\mathbf{g}_{[t]} + \boldsymbol{\lambda}_{[t]}^{-1}\mathbf{1}\Big],$$

then, we get

$$\begin{aligned}
&\left\|\tilde{\boldsymbol{\lambda}}_{[t+1]}\right\| \\
&= \Big\|\Big[\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \times \left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right) - \boldsymbol{\lambda}_{[t]}^{-1}\mathbf{Q}_{[t]}\Big]^{-1} \\
&\quad \times \Big[-\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\mathbf{g}_{[t]} + \boldsymbol{\lambda}_{[t]}^{-1}\mathbf{1}\Big]\Big\| \\
&\le \Big\|\Big[\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right)\Big]^{-1} \\
&\quad \times \Big[-\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\mathbf{g}_{[t]} + \boldsymbol{\lambda}_{[t]}^{-1}\mathbf{1}\Big]\Big\| \\
&\le \boldsymbol{\lambda}_{\min}^{-1}\Big\{\Big[\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \times \left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right)\Big]\Big\} \times \Big(\Big\|\Big[-\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right) \\
&\quad \times \left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\mathbf{g}_{[t]}\Big]\Big\| + \left\|\boldsymbol{\lambda}_{[t]}^{-1}\mathbf{1}\right\|\Big) \\
&\le \boldsymbol{\lambda}_{\min}^{-1}\Big\{\Big[\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \times \left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right)\Big]\Big\} \times \Big(\boldsymbol{\lambda}_{\min}^{-1}\left\{\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)\right\} \\
&\quad \times \left\|\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\mathbf{g}_{[t]}\right\| + \left\|\boldsymbol{\lambda}_{[t]}^{-1}\mathbf{1}\right\|\Big),
\end{aligned}$$

where $\boldsymbol{\lambda}_{\min}\Big\{\Big[\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^{\top}-\nabla\mathbf{e}_{[t]}\right)\Big]\Big\}$ denotes the smallest eigenvalue of $\Big[\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^{\top}-\nabla\mathbf{e}_{[t]}\right)\Big]$, and $\boldsymbol{\lambda}_{\min}\left\{\left(H_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)\right\}$ denotes the smallest eigenvalue of $(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]})$. The first inequality holds because $-\boldsymbol{\lambda}_{[t]}^{-1}\mathbf{Q}_{[t]}$ is a positive definite diagonal matrix (as a result of the strict feasibility of $\mathbf{y}_{[t]}$ and $\boldsymbol{\lambda}_{[t]}$) and because its eigenvalues are all positive, which will increase the eigenvalues of $\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right)$. According to the norm triangle inequality, the second and third inequalities also hold.

Because $\mathbf{g}_{[t]}$ is continuous, the spectral radius $\rho\left(\mathbf{H}_{[t]}\right)$ is bounded. Similarly, because the elements in $\mathbf{y}_{[t]}$ are all positive numbers and bounded, the elements in $\nabla\mathbf{e}_{[t]}$ and diagonal matrix $-\nabla^2\mathbf{e}_{[t]}$ are all bounded nonnegative numbers. Matrix $\mathbf{M}$ is a constant matrix representing network topology information. Hence, $\boldsymbol{\lambda}_{\min}^{-1}\Big\{\Big[\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^{\top}\right)\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^{\top} - \nabla\mathbf{e}_{[t]}\right)\Big]\Big\}$ must be finite. At the same time, due to the projection on a set, $\mathbf{y}_{[t]}$ and $\boldsymbol{\lambda}_{[t]}$ are strictly bounded away from zero, so $\|\mathbf{g}_{[t]}\|$ and $\left\|\boldsymbol{\lambda}_{[t]}^{-1}\mathbf{1}\right\|$ are bounded. Hence, $\left\|\tilde{\boldsymbol{\lambda}}_{[t+1]}\right\| < \infty$ holds. Finally, note that, according to the norm triangle inequality, $\|\boldsymbol{\lambda}_{[t+1]}\| = \left\|(1-\pi)\,\boldsymbol{\lambda}_{[t]} + \pi\tilde{\boldsymbol{\lambda}}_{[t+1]}\right\| \le (1-\pi)\,\|\boldsymbol{\lambda}_{[t]}\| + \pi\left\|\tilde{\boldsymbol{\lambda}}_{[t+1]}\right\|$. In summary, we conclude that $\|\boldsymbol{\lambda}_{[t+1]}\| < \infty$. This completes the proof of Theorem 2.

In order to prove the convergence of JCCPA, we need to analyze the Lyapunov drift rate of a time slot.

The Lyapunov drift function is defined as follows:

$$V\left(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}\right) = \frac{1}{2\pi}\left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|^2 + \frac{1}{2\mu^3\pi}\left\|\boldsymbol{\lambda}_{[t]} - \boldsymbol{\lambda}^*\right\|^2,$$

which represents the distance between the primal-dual iterate $(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]})$ and the perturbed KKT point $(\mathbf{y}^*, \boldsymbol{\lambda}^*)$. Therefore, the one-slot Lyapunov drift $\Delta V\left(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}\right)$ can be described as follows:

$$\begin{aligned}
&\Delta V\left(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}\right) \\
&\quad = V\left(\mathbf{y}_{[t+1]}, \boldsymbol{\lambda}_{[t+1]}\right) - V\left(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}\right) \\
&\quad = \frac{1}{2\pi}(\mathbf{y}_{[t+1]} + \mathbf{y}_{[t]} - 2\mathbf{y}^*)^\top(\mathbf{y}_{[t+1]} - \mathbf{y}_{[t]}) \qquad (38)\\
&\qquad + \frac{1}{2\mu^3\pi}(\boldsymbol{\lambda}_{[t+1]} + \boldsymbol{\lambda}_{[t]} - 2\boldsymbol{\lambda}^*)^\top(\boldsymbol{\lambda}_{[t+1]} - \boldsymbol{\lambda}_{[t]}). \qquad (39)
\end{aligned}$$

Next, we discuss (38) and (39), respectively.

A. DISCUSSION OF EQUATION (38)

$$\begin{aligned}
(38) &= \Big[\frac{1}{\pi}\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right) - \frac{1}{2}\mathbf{F}_{[t]}^{-1}\Big[\mathbf{g}_{[t]} \\
&\quad - \left(\mathbf{M}^\top - \nabla \mathbf{e}_{[t]}\right)\mathbf{Q}_{[t]}^{-1}\mathbf{1}\Big]\Big]^\top \\
&\quad \times \Big[-\pi\mathbf{F}_{[t]}^{-1}\Big[\mathbf{g}_{[t]} - \left(\mathbf{M}^\top - \nabla \mathbf{e}_{[t]}\right)\mathbf{Q}_{[t]}^{-1}\mathbf{1}\Big]\Big] \\
&= -\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)^\top \mathbf{F}_{[t]}^{-1} \\
&\quad \times \Big[\mathbf{g}_{[t]} - \left(\mathbf{M}^\top - \nabla \mathbf{e}_{[t]}\right)\mathbf{Q}_{[t]}^{-1}\mathbf{1}\Big] \qquad (40)\\
&\quad + \frac{\pi}{2}\Big[\mathbf{g}_{[t]} - \left(\mathbf{M}^\top - \nabla \mathbf{e}_{[t]}\right)\mathbf{Q}_{[t]}^{-1}\mathbf{1}\Big]^\top \\
&\quad \times \left(\mathbf{F}_{[t]}^{-1}\right)^2\Big[\mathbf{g}_{[t]} - \left(\mathbf{M}^\top - \nabla \mathbf{e}_{[t]}\right)\mathbf{Q}_{[t]}^{-1}\mathbf{1}\Big]. \qquad (41)
\end{aligned}$$

Based on perturbed KKT conditions (11) and (14), we have: $\mathbf{g}^* + (\mathbf{M}^\top - \nabla \mathbf{e}^*)\boldsymbol{\lambda}^* = \mathbf{0}$, where $\boldsymbol{\lambda}^* = -\left(\mathbf{Q}^*\right)^{-1}\mathbf{1}$. So, (40) can be converted into

$$\begin{aligned}
(40) &= -\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)^\top \mathbf{F}_{[t]}^{-1}\Big[\mathbf{g}_{[t]} - \mathbf{g}^* - \left(\mathbf{M}^\top - \nabla \mathbf{e}^*\right)\boldsymbol{\lambda}^* \\
&\quad - \left(\mathbf{M}^\top - \nabla \mathbf{e}_{[t]}\right)\mathbf{Q}_{[t]}^{-1}\mathbf{1}\Big] \\
&= -\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)^\top \mathbf{F}_{[t]}^{-1}\Big[\mathbf{g}_{[t]} - \mathbf{g}^* \\
&\quad + \left(\mathbf{M}^\top - \nabla \mathbf{e}^*\right)\mathbf{Q}_*^{-1}\mathbf{1} - \left(\mathbf{M}^\top - \nabla \mathbf{e}_{[t]}\right)\mathbf{Q}_{[t]}^{-1}\mathbf{1}\Big] \\
&= -\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)^\top \mathbf{F}_{[t]}^{-1}\left(\mathbf{g}_{[t]} - \mathbf{g}^*\right) \qquad (42)\\
&\quad + \left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)^\top \mathbf{F}_{[t]}^{-1}\Big[\left(\mathbf{M}^\top - \nabla \mathbf{e}_{[t]}\right)\mathbf{Q}_{[t]}^{-1} \\
&\quad - \left(\mathbf{M}^\top - \nabla \mathbf{e}^*\right)\left(\mathbf{Q}^*\right)^{-1}\Big]\mathbf{1}. \qquad (43)
\end{aligned}$$

For (42), we have

$$\begin{aligned}
(42) &= -\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)^\top \mathbf{F}_{[t]}^{-1}\left(\mathbf{g}_{[t]} - \mathbf{g}^*\right) \\
&\le -\frac{1}{\boldsymbol{\lambda}_{\min}\{\mathbf{F}\}}\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)^\top\left(\mathbf{g}_{[t]} - \mathbf{g}^*\right), \qquad (44)
\end{aligned}$$

where $\boldsymbol{\lambda}_{\min}\{\mathbf{F}\} = \inf_t\left\{\boldsymbol{\lambda}_{\min}\left\{\mathbf{F}_{[t]}\right\}\right\}$ denotes the lower bound of the smallest eigenvalue in all time slots of $\mathbf{F}_{[t]}$. By analogy, the eigenvalue of other matrices can be defined in the same way.

In order to obtain the bound of $\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)^\top\left(\mathbf{g}_{[t]} - \mathbf{g}^*\right)$, we first obtain the following results according to the Taylor Mean-Value Theorem:

$$\begin{aligned}
f_\mu\left(\mathbf{y}_{[t]}\right) &= f_\mu\left(\mathbf{y}^*\right) + \left(\mathbf{g}^*\right)^\top\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right) \\
&\quad + \frac{1}{2}\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)^\top H[\tilde{\mathbf{y}}_1]\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right), \\
f_\mu\left(\mathbf{y}^*\right) &= f_\mu\left(\mathbf{y}_{[t]}\right) + \left(\mathbf{g}_{[t]}\right)^\top\left(\mathbf{y}^* - \mathbf{y}_{[t]}\right) \\
&\quad + \frac{1}{2}\left(\mathbf{y}^* - \mathbf{y}_{[t]}\right)^\top \mathbf{H}[\tilde{\mathbf{y}}_2]\left(\mathbf{y}^* - \mathbf{y}_{[t]}\right),
\end{aligned}$$

where $\mathbf{H}[\tilde{\mathbf{y}}_1]$ and $\mathbf{H}[\tilde{\mathbf{y}}_2]$ represent the Hessian matrices at points $\tilde{\mathbf{y}}_1$ and $\tilde{\mathbf{y}}_2$, $\tilde{\mathbf{y}}_1 = (1-\alpha_1)\,\mathbf{y}_{[t]} + \alpha_1\mathbf{y}^*$, and $\tilde{\mathbf{y}}_2 = (1-\alpha_2)\,\mathbf{y}_{[t]} + \alpha_2\mathbf{y}^*$, for $0 \le \alpha_1, \alpha_2 \le 1$. Based on the above equations, we can obtain

$$\begin{aligned}
&\left(\mathbf{g}_{[t]} - \mathbf{g}^*\right)^\top\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right) \\
&\quad = \frac{1}{2}\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)^\top\left(\mathbf{H}[\tilde{\mathbf{y}}_1] + \mathbf{H}[\tilde{\mathbf{y}}_2]\right)\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right) \\
&\quad \ge \boldsymbol{\lambda}_{\min}\{\mathbf{H}\}\left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|^2.
\end{aligned}$$

Therefore, by combining the above equation with (44), the following result can be introduced

$$(42) \le -\mathbf{B}_1\left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|^2, \qquad (45)$$

where $\mathbf{B}_1 = \frac{\lambda_{\min}\{\mathbf{H}\}}{\lambda_{\min}\{\mathbf{F}\}}$.

In the definition of $\mathbf{F}_{[t]}$: $\mathbf{F}_{[t]} = \Big[\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]} - \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)\mathbf{Q}_{[t]}^{-1}\boldsymbol{\lambda}_{[t]}\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right)\Big]$, $\mathbf{H}_{[t]}$, $\nabla^2\mathbf{e}_{[t]}$, and $\mathbf{Q}_{[t]}^{-1}\boldsymbol{\lambda}_{[t]}$ are all diagonal matrices. Some diagonal elements in $\mathbf{H}_{[t]}$ tend to be infinite with $\mu \to \infty$, and the rest are independent of μ. $\nabla^2\mathbf{e}_{[t]}$, $\mathbf{Q}_{[t]}^{-1}\boldsymbol{\lambda}_{[t]}$, and $\left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)$ are independent of μ, so both $\boldsymbol{\lambda}_{\min}\{\mathbf{H}\}$ and $\boldsymbol{\lambda}_{\min}\{\mathbf{F}\}$ are independent of μ. That is to say, the value of $\mathbf{B}_1$ is also independent of μ.

We now discuss the term in (43). The numerator and denominator are multiplied by $\mathbf{Q}_{[t]}\mathbf{Q}^*$ simultaneously, so we have

$$\begin{aligned}
(43) &\le \frac{1}{\boldsymbol{\lambda}_{\min}\{\mathbf{F}\}\Gamma_1}\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)^\top \\
&\quad \times \Big[\left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)\mathbf{Q}^* - \left(\mathbf{M}^\top - \nabla\mathbf{e}^*\right)\mathbf{Q}_{[t]}\Big]\mathbf{1}, \qquad (46)
\end{aligned}$$

where $\Gamma_1 = \inf_t \lambda_{\min}\left\{\mathbf{Q}_{[t]}\mathbf{Q}^*\right\}$.

Due to $\mathbf{Q}_{[t]}\mathbf{1} = \mathbf{M}\mathbf{y}_{[t]} - \mathbf{e}_{[t]}$, we have the following relationships according to the vectorial Taylor expansion:

$$\begin{aligned}
\mathbf{Q}_{[t]}\mathbf{1} &= \mathbf{Q}^*\mathbf{1} + \left(\mathbf{M}^\top - \nabla\mathbf{e}^*\right)^\top\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right) \\
&\quad + o\left(\left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|\right)\mathbf{1}, \\
\mathbf{Q}^*\mathbf{1} &= \mathbf{Q}_{[t]}\mathbf{1} + \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)^\top\left(\mathbf{y}^* - \mathbf{y}_{[t]}\right) \\
&\quad + o\left(\left\|\mathbf{y}^* - \mathbf{y}_{[t]}\right\|\right)\mathbf{1}.
\end{aligned}$$

After shifting items, we obtain

$$\begin{aligned}
&\mathbf{Q}_{[t]}\mathbf{1} - \mathbf{Q}^*\mathbf{1} - o\left(\left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|\right)\mathbf{1} \\
&\quad = \left(\mathbf{M}^\top - \nabla\mathbf{e}^*\right)^\top\left(\mathbf{y}_{[t]} - \mathbf{y}^*\right), \\
&\mathbf{Q}^*\mathbf{1} - \mathbf{Q}_{[t]}\mathbf{1} - o\left(\left\|\mathbf{y}^* - \mathbf{y}_{[t]}\right\|\right)\mathbf{1} \\
&\quad = \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)^\top\left(\mathbf{y}^* - \mathbf{y}_{[t]}\right).
\end{aligned}$$

Substitute them into (46) and simplify it as follows:

$$\begin{aligned}(46) &= \frac{1}{\lambda_{\min}\{\mathbf{F}\}\Gamma_1}\Big[\Big(\mathbf{Q}_{[t]}\mathbf{1}-\mathbf{Q}^*\mathbf{1} \\ &\quad + o\left(\|\mathbf{y}^*-\mathbf{y}_{[t]}\|\right)\mathbf{1}\Big)^\top \mathbf{Q}^*\mathbf{1} - \Big(\mathbf{Q}_{[t]}\mathbf{1}-\mathbf{Q}^*\mathbf{1} \\ &\quad - o\left(\|\mathbf{y}_{[t]}-\mathbf{y}^*\|\right)\mathbf{1}\Big)^\top \mathbf{Q}_{[t]}\mathbf{1}\Big] \\ &= -\frac{1}{\lambda_{\min}\{\mathbf{F}\}\Gamma_1}\Big[\left\|\mathbf{Q}_{[t]}\mathbf{1}-\mathbf{Q}^*\mathbf{1}\right\|^2 \\ &\quad + \left[o\left(\|\mathbf{y}_{[t]}-\mathbf{y}^*\|\right)\mathbf{1}\right]^\top\left(\mathbf{Q}^*\mathbf{1}+\mathbf{Q}_{[t]}\mathbf{1}\right)\Big] \\ &\le 0. \end{aligned} \tag{47}$$

Hence, combining (45) and (47), we have

$$(40) \le -\mathbf{B}_1\left\|\mathbf{y}_{[t]}-\mathbf{y}^*\right\|^2. \tag{48}$$

The upper bound of (38) has been found. Next, we discuss the upper bound of (41).

According to the norm triangle inequality and the perturbed KKT conditions (11) and (14): $\mathbf{g}^*+(\mathbf{M}^\top-\nabla\mathbf{e}^*)\boldsymbol{\lambda}^*=\mathbf{0}$, $\boldsymbol{\lambda}^*=-\left(\mathbf{Q}^*\right)^{-1}\mathbf{1}$, we have

$$\begin{aligned}(41) &\le \frac{\pi}{2\lambda^2_{\min}\{\mathbf{F}\}}\left\|\mathbf{g}_{[t]}-\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\mathbf{Q}^{-1}_{[t]}\mathbf{1}\right\|^2 \\ &= \frac{\pi}{2\lambda^2_{\min}\{\mathbf{F}\}}\Big\|\mathbf{g}_{[t]}-\mathbf{g}^*-\left(\mathbf{M}^\top-\nabla\mathbf{e}^*\right)\boldsymbol{\lambda}^* \\ &\quad -\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\mathbf{Q}^{-1}_{[t]}\mathbf{1}\Big\|^2 \\ &= \frac{\pi}{2\lambda^2_{\min}\{\mathbf{F}\}}\Big\|\mathbf{g}_{[t]}-\mathbf{g}^*+\left(\mathbf{M}^\top-\nabla\mathbf{e}^*\right)\left(\mathbf{Q}^*\right)^{-1}\mathbf{1} \\ &\quad -\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\mathbf{Q}^{-1}_{[t]}\mathbf{1}\Big\|^2 \\ &= \frac{\pi}{2\lambda^2_{\min}\{\mathbf{F}\}}\Big\|\mathbf{g}_{[t]}-\mathbf{g}^*-\mathbf{M}^\top\left(\mathbf{Q}^{-1}_{[t]}-\left(\mathbf{Q}^*\right)^{-1}\right)\mathbf{1} \\ &\quad +\nabla\mathbf{e}_{[t]}\mathbf{Q}^{-1}_{[t]}\mathbf{1}-\nabla\mathbf{e}^*\left(\mathbf{Q}^*\right)^{-1}\mathbf{1}\Big\|^2 \\ &\le \frac{\pi}{2\lambda^2_{\min}\{\mathbf{F}\}}\Big[\left\|\mathbf{g}_{[t]}-\mathbf{g}^*\right\|^2 \\ &\quad +\left\|\mathbf{M}^\top\left(\mathbf{Q}^{-1}_{[t]}-\left(\mathbf{Q}^*\right)^{-1}\right)\mathbf{1}\right\|^2 \\ &\quad +\left\|\nabla\mathbf{e}_{[t]}\mathbf{Q}^{-1}_{[t]}\mathbf{1}-\nabla\mathbf{e}^*\left(\mathbf{Q}^*\right)^{-1}\mathbf{1}\right\|^2\Big]. \end{aligned} \tag{49}$$

Note that the μ-factors in (49) can be cancelled mutually. Moreover, because the utility function and channel capacity function are differentiable, and $\mathbf{y}_{[t]}$ is bounded, we can conclude that (49) is upper-bounded. Let $\mathbf{B}_2 = \frac{1}{2\lambda^2_{\min}\{\mathbf{F}\}}\sup_t\left\{\left[\left\|\mathbf{g}_{[t]}-\mathbf{g}^*\right\|^2+\left\|\mathbf{M}^\top(\mathbf{Q}^{-1}_{[t]}-(\mathbf{Q}^*)^{-1})\mathbf{1}\right\|^2+\left\|\nabla\mathbf{e}_{[t]}\mathbf{Q}^{-1}_{[t]}\mathbf{1}-\nabla\mathbf{e}^*\left(\mathbf{Q}^*\right)^{-1}\mathbf{1}\right\|^2\right]\right\}$ denote the upper bound of the part of (49) except π. Combining (48) with (49), we have

$$(38) = (40)+(41) \le -\mathbf{B}_1\left\|\mathbf{y}_{[t]}-\mathbf{y}^*\right\|^2+\pi\mathbf{B}_2. \tag{50}$$

Thus far, we have completed the analysis for (38).

B. DISCUSSION OF EQUATION (39)

$$\begin{aligned}&(39) \\ &= \frac{1}{2\mu^3\pi}\left(\boldsymbol{\lambda}_{[t+1]}+\boldsymbol{\lambda}_{[t]}-2\boldsymbol{\lambda}^*\right)^\top\left(\boldsymbol{\lambda}_{[t+1]}-\boldsymbol{\lambda}_{[t]}\right) \\ &= \frac{1}{2\mu^3\pi}\Big[2\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right) \\ &\quad -\pi\mathbf{G}^{-1}_{[t]}\Big[\left(\mathbf{M}-\nabla\mathbf{e}^\top_{[t]}\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1} \\ &\quad \times\left(\mathbf{g}_{[t]}+\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}_{[t]}\right)-\left(\mathbf{Q}_{[t]}+\boldsymbol{\lambda}^{-1}_{[t]}\right)\mathbf{1}\Big]\Big]^\top \\ &\quad \times\Big[-\pi\mathbf{G}^{-1}_{[t]}\Big[\left(\mathbf{M}-\nabla\mathbf{e}^\top_{[t]}\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1} \\ &\quad \times\left(\mathbf{g}_{[t]}+\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}_{[t]}\right)-\left(\mathbf{Q}_{[t]}+\boldsymbol{\lambda}^{-1}_{[t]}\right)\mathbf{1}\Big]\Big] \\ &= -\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top\times\mathbf{G}^{-1}_{[t]}\Big[\left(\mathbf{M}-\nabla\mathbf{e}^\top_{[t]}\right) \\ &\quad \times\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{g}_{[t]}+\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}_{[t]}\right) \\ &\quad -(\mathbf{Q}_{[t]}+\boldsymbol{\lambda}^{-1}_{[t]})\mathbf{1}\Big] \qquad (51) \\ &\quad +\frac{\pi}{2\mu^3}\Big[\left(\mathbf{M}-\nabla\mathbf{e}^\top_{[t]}\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1} \\ &\quad \times\left(\mathbf{g}_{[t]}+\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}_{[t]}\right)-(\mathbf{Q}_{[t]}+\boldsymbol{\lambda}^{-1}_{[t]})\mathbf{1}\Big]^\top \\ &\quad \times\left(\mathbf{G}^{-1}_{[t]}\right)^2\times\Big[\left(\mathbf{M}-\nabla\mathbf{e}^\top_{[t]}\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1} \\ &\quad \times\left(\mathbf{g}_{[t]}+\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}_{[t]}\right)-(\mathbf{Q}_{[t]}+\boldsymbol{\lambda}^{-1}_{[t]})\mathbf{1}\Big]. \end{aligned} \tag{52}$$

Note that, according to (19), $\mathbf{H}_{[t]}$ is in the expression of $\mathbf{G}_{[t]}$, so the bound of $\mathbf{G}^{-1}_{[t]}$ has relationship with μ.

First, according to perturbed KKT conditions (11): $\mathbf{g}^*+(\mathbf{M}^\top-\nabla\mathbf{e}^*)\boldsymbol{\lambda}^*=\mathbf{0}$, (51) can be further decomposed as follows:

$$\begin{aligned}(51) &= -\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top\mathbf{G}^{-1}_{[t]}\Big[\left(\mathbf{M}-\nabla\mathbf{e}^\top_{[t]}\right) \\ &\quad \times\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\times\Big[\mathbf{g}_{[t]}+\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}_{[t]} \\ &\quad -\mathbf{g}^*-\left(\mathbf{M}^\top-\nabla\mathbf{e}^*\right)\boldsymbol{\lambda}^*\Big]-\left(\mathbf{Q}_{[t]}+\boldsymbol{\Lambda}^{-1}_{[t]}\right)\mathbf{1}\Big] \\ &= -\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top\mathbf{G}^{-1}_{[t]}\left(\mathbf{M}-\nabla\mathbf{e}^\top_{[t]}\right) \\ &\quad \times\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\Big[\left(\mathbf{g}_{[t]}-\mathbf{g}^*\right) \\ &\quad +\left(\nabla\mathbf{e}^*-\nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}^*\Big] \\ &\quad -\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top\mathbf{G}^{-1}_{[t]}\left(\mathbf{M}-\nabla\mathbf{e}^\top_{[t]}\right) \\ &\quad \times\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\Big[\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\boldsymbol{\Lambda}_{[t]} \\ &\quad -\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}^*\Big] \\ &\quad +\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top\mathbf{G}^{-1}_{[t]}\left(\mathbf{Q}_{[t]}+\boldsymbol{\Lambda}^{-1}_{[t]}\right)\mathbf{1} \end{aligned}$$

$$\le -\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\times\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\left[\left(\mathbf{g}_{[t]}-\mathbf{g}^*\right)+\left(\nabla\mathbf{e}^*-\nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}^*\right] \quad (53)$$

$$+\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\left(\mathbf{Q}_{[t]}+\boldsymbol{\Lambda}_{[t]}^{-1}\right)\mathbf{1}. \quad (54)$$

The inequality holds, because $\mathbf{G}_{[t]}^{-1}\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)$ is a semipositive definite matrix, which means

$$-\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\times\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)\le 0.$$

For (54), according to perturbed KKT conditions (14): $\boldsymbol{\lambda}^*=-\left(\mathbf{Q}^*\right)^{-1}\mathbf{1}$, i.e., $(\mathbf{Q}^*-\left(\boldsymbol{\Lambda}^*\right)^{-1})\mathbf{1}=\mathbf{0}$, we can further obtain

$$(54)=\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\times\left(\mathbf{Q}_{[t]}-\mathbf{Q}^*-\left(\boldsymbol{\Lambda}^*\right)^{-1}+\boldsymbol{\Lambda}_{[t]}^{-1}\right)\mathbf{1}$$
$$=\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\mathbf{M}\left(\mathbf{y}_{[t]}-\mathbf{y}^*\right)-\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\left(\mathbf{e}_{[t]}-\mathbf{e}^*\right)+\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\left[\left(\boldsymbol{\Lambda}_{[t]}^{-1}-\left(\boldsymbol{\Lambda}^*\right)^{-1}\right)\mathbf{1}\right]$$
$$\le\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top\times\mathbf{G}_{[t]}^{-1}\left[\left(\mathbf{M}\mathbf{y}_{[t]}-\mathbf{e}_{[t]}\right)-\left(\mathbf{M}\mathbf{y}^*-\mathbf{e}^*\right)\right]. \quad (55)$$

The inequality holds because

$$\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\left[(\boldsymbol{\Lambda}_{[t]}^{-1}-\left(\boldsymbol{\Lambda}^*\right)^{-1})\mathbf{1}\right]\le-\frac{1}{\mu^3}\frac{1}{\boldsymbol{\lambda}_{\min}\{\mathbf{G}\}\Gamma_2}\left\|\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)\right\|^2,$$

where $\Gamma_2=\inf_t\left\{\boldsymbol{\lambda}_{\min}\left\{\boldsymbol{\Lambda}_{[t]}\boldsymbol{\Lambda}^*\right\}\right\}$.

Further transform (55) into the following equation:

$$(55)=\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\times\left[\left(\mathbf{M}\mathbf{y}_{[t]}-\mathbf{e}_{[t]}\right)-\left(\mathbf{M}\mathbf{y}^*-\mathbf{e}^*\right)\right]$$
$$=\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\left[\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)^\top\times\left(\mathbf{y}_{[t]}-\mathbf{y}^*\right)-o\left(\left\|\mathbf{y}^*-\mathbf{y}_{[t]}\right\|\right)\mathbf{1}\right]$$
$$\le\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)^\top\left(\mathbf{y}_{[t]}-\mathbf{y}^*\right). \quad (56)$$

Combining (53) with (56), we have

$$(51)\le-\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\times\left[\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\left[\left(\mathbf{g}_{[t]}-\mathbf{g}^*\right)+(\nabla\mathbf{e}^*-\nabla\mathbf{e}_{[t]})\boldsymbol{\lambda}^*\right]-\left(\mathbf{y}_{[t]}-\mathbf{y}^*\right)\right]. \quad (57)$$

According to the vectorial Taylor expansion for multivariate functions, $\mathbf{g}^*-\nabla\mathbf{e}^*\boldsymbol{\lambda}^*$ can be expanded at points $(\mathbf{y}_{[t]},\boldsymbol{\lambda}_{[t]})$ as follows:

$$\mathbf{g}^*-\nabla\mathbf{e}^*\boldsymbol{\lambda}^*=\mathbf{g}_{[t]}-\nabla\mathbf{e}_{[t]}\boldsymbol{\lambda}_{[t]}+\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)\left(\mathbf{y}^*-\mathbf{y}_{[t]}\right)+\left(-\nabla\mathbf{e}_{[t]}\right)\left(\boldsymbol{\lambda}^*-\boldsymbol{\lambda}_{[t]}\right)+o\left(\rho\right)\mathbf{1}$$
$$=\mathbf{g}_{[t]}-\nabla\mathbf{e}_{[t]}\boldsymbol{\lambda}^*+\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)\left(\mathbf{y}^*-\mathbf{y}_{[t]}\right)+o\left(\rho\right)\mathbf{1}, \quad (58)$$

where $\rho=\sqrt{\left\|\mathbf{y}_*-\mathbf{y}_{[t]}\right\|^2+\left\|\boldsymbol{\lambda}^*-\boldsymbol{\lambda}_{[t]}\right\|^2}$.

Combining (57) with (58), and according to the Cauchy-Schwarz inequality, we have

$$(51)\le\frac{1}{\mu^3}\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)^\top \mathbf{G}_{[t]}^{-1}\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\rho\mathbf{1}\le\frac{\rho\left\|\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right\|\left\|\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\mathbf{1}\right\|}{\mu^3\boldsymbol{\lambda}_{\min}\{\mathbf{G}\}}. \quad (59)$$

Because of the boundedness of $\mathbf{y}_{[t]}$, $\nabla\mathbf{e}_{[t]}^\top$ is bounded. $\mathbf{M}$ is a constant matrix representing network topology information, so $\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\mathbf{1}$ is bounded. $\boldsymbol{\lambda}_{[t]}$ is bounded according to Lemma 2. Hence, the inequality (59) holds.

Let $\mathbf{B}_3=\frac{1}{\mu^2\boldsymbol{\lambda}_{\min}\{\mathbf{G}\}}\sup_t\left\{\rho\left\|\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right\|\left\|\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\mathbf{1}\right\|\right\}$, and we have

$$(51)\le\frac{1}{\mu}\mathbf{B}_3. \quad (60)$$

Finally, we discuss the upper bound of (52). In the step control strategy, $\pi\in(0,1]$, so we have

$$(52)\le\frac{1}{2\mu^3}\left[\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\times\left(\mathbf{g}_{[t]}+(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]})\boldsymbol{\lambda}_{[t]}\right)-\left(\mathbf{Q}_{[t]}+\boldsymbol{\Lambda}_{[t]}^{-1}\right)\mathbf{1}\right]^\top\times\left(\mathbf{G}_{[t]}^{-1}\right)^2\left[\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\times\left(\mathbf{g}_{[t]}+(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]})\boldsymbol{\lambda}_{[t]}\right)-\left(\mathbf{Q}_{[t]}+\boldsymbol{\Lambda}_{[t]}^{-1}\right)\mathbf{1}\right]$$
$$\le\frac{1}{2\mu^3\boldsymbol{\lambda}_{\min}^2\{\mathbf{G}\}}\left\|\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\times\left(\mathbf{g}_{[t]}+(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]})\boldsymbol{\lambda}_{[t]}\right)-\left(\mathbf{Q}_{[t]}+\boldsymbol{\Lambda}_{[t]}^{-1}\right)\mathbf{1}\right\|^2. \quad (61)$$

According to the perturbed KKT conditions (14): $\boldsymbol{\lambda}^*=-\left(\mathbf{Q}^*\right)^{-1}\mathbf{1}$ and (11): $\mathbf{g}^*+(\mathbf{M}^\top-\nabla\mathbf{e}^*)\boldsymbol{\lambda}^*=\mathbf{0}$, referring to the expression of $\mathbf{Q}^*\mathbf{1}-\mathbf{Q}_{[t]}\mathbf{1}$ obtained above, (61) can be transformed into

$$(61)=\frac{1}{2\mu^3\boldsymbol{\lambda}_{\min}^2\{\mathbf{G}\}}\left\|\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\times\left(\mathbf{g}_{[t]}+\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}_{[t]}\right)-\left(\mathbf{Q}_{[t]}-\mathbf{Q}^*+\boldsymbol{\Lambda}_{[t]}^{-1}-\left(\boldsymbol{\Lambda}^*\right)^{-1}\right)\mathbf{1}\right\|^2$$

$$
\begin{aligned}
&= \frac{1}{2\mu^3\boldsymbol{\lambda}_{\min}^2\{\mathbf{G}\}} \Big\| \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) \\
&\quad \times\Big[\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \left(\mathbf{g}_{[t]} + \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}_{[t]}\right) \\
&\quad - \left(\mathbf{y}_{[t]} - \mathbf{y}^*\right) \Big] \\
&\quad + \left(\left\|\mathbf{y}^* - \mathbf{y}_{[t]}\right\|\right)\mathbf{1} - \left(\boldsymbol{\Lambda}_{[t]}^{-1} - \boldsymbol{\Lambda}_*^{-1}\right)\mathbf{1} \Big\|^2 \\
&= \frac{1}{2\mu^3\boldsymbol{\Lambda}_{\min}^2\{\mathbf{G}\}} \Big\| \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) \Big[\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \\
&\quad \times \left((\mathbf{g}_{[t]} - \mathbf{g}^*) + (\nabla\mathbf{e}^* - \nabla\mathbf{e}_{[t]})\,\boldsymbol{\lambda}^*\right) - \left(\mathbf{y}_{[t]} - \mathbf{y}^*\right) \Big] \\
&\quad + \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) \Big[\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \\
&\quad \times\Big(\mathbf{g}^* + \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right) \\
&\quad \times\boldsymbol{\lambda}_{[t]} - (\nabla\mathbf{e}^* - \nabla\mathbf{e}_{[t]})\,\boldsymbol{\lambda}^*\Big)\Big] \\
&\quad + \left(\left\|\mathbf{y}^* - \mathbf{y}_{[t]}\right\|\right)\mathbf{1} - (\boldsymbol{\Lambda}_{[t]}^{-1} - (\boldsymbol{\Lambda}^*)^{-1})\mathbf{1} \Big\|^2 \\
&= \frac{1}{2\mu^3\boldsymbol{\lambda}_{\min}^2\{\mathbf{G}\}} \Big\| \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) \\
&\quad \times\Big[\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \Big((\mathbf{g}_{[t]} - \mathbf{g}^*) \\
&\quad + (\nabla\mathbf{e}^* - \nabla\mathbf{e}_{[t]})\,\boldsymbol{\lambda}^*\Big) - \left(\mathbf{y}_{[t]} - \mathbf{y}^*\right) \Big] \\
&\quad + \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) \left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \\
&\quad \times \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right) \left(\boldsymbol{\lambda}_{[t]} - \boldsymbol{\lambda}^*\right) \\
&\quad + \left(\left\|\mathbf{y}^* - \mathbf{y}_{[t]}\right\|\right)\mathbf{1} - (\boldsymbol{\Lambda}_{[t]}^{-1} - (\boldsymbol{\Lambda}^*)^{-1})\mathbf{1} \Big\|^2. \qquad (62)
\end{aligned}
$$

According to the norm triangle inequality, we have

$$
\begin{aligned}
(62) \le &\frac{1}{2\mu^3\boldsymbol{\lambda}_{\min}^2\{\mathbf{G}\}} \Big\{ \Big\| \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) \Big[\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \\
&\times \left[(\mathbf{g}_{[t]} - \mathbf{g}^*) + (\nabla\mathbf{e}^* - \nabla\mathbf{e}_{[t]})\,\boldsymbol{\lambda}^*\right] - (\mathbf{y}_{[t]} - \mathbf{y}^*)\Big]\Big\| \\
&+ \Big\| \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) \left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right) \\
&\times \left(\boldsymbol{\lambda}_{[t]} - \boldsymbol{\lambda}^*\right) \Big\| + \left\|\mathbf{y}^* - \mathbf{y}_{[t]}\right\| \\
&+ \left\|(\boldsymbol{\lambda}_{[t]}^{-1} - (\boldsymbol{\lambda}^*)^{-1})\mathbf{1}\right\| \Big\}^2. \qquad (63)
\end{aligned}
$$

Note that the first part $\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right)\Big[\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}$ $\Big[\left(\mathbf{g}_{[t]} - \mathbf{g}^*\right) + \left(\nabla\mathbf{e}^* - \nabla\mathbf{e}_{[t]}\right)\boldsymbol{\lambda}^*\Big] - \left(\mathbf{y}_{[t]} - \mathbf{y}^*\right)\Big]$ in (63) has the same form as (57), so

$$
\begin{aligned}
&\Big\| \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) \Big[\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \Big[(\mathbf{g}_{[t]} - \mathbf{g}^*) \\
&\quad + (\nabla\mathbf{e}^* - \nabla\mathbf{e}_{[t]})\,\boldsymbol{\lambda}^*\Big] - (\mathbf{y}_{[t]} - \mathbf{y}^*) \Big]\Big\| \\
&\quad \le \rho \left\|\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right)\mathbf{1}\right\|,
\end{aligned}
$$

and $\rho \left\|\left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right)\mathbf{1}\right\|$ has upper bound.

For the second part $\left\|\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)\left(\boldsymbol{\lambda}_{[t]} - \boldsymbol{\lambda}^*\right)\right\|$ in (63), the μ factor in $\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1}$ is cancelled with that in $\boldsymbol{\lambda}_{[t]} - \boldsymbol{\lambda}^*$. Moreover, $\left\|\boldsymbol{\lambda}_{[t]}\right\|$ is bounded according to Lemma 2. In matrix $\left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)$, $\mathbf{M}$ is a constant matrix representing network topology information and $\nabla\mathbf{e}_{[t]}$ is bounded due to the boundedness of $\mathbf{y}_{[t]}$, so $\left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right)$is bounded. Based on these points, the second part $\left\|\left(\mathbf{M}-\nabla\mathbf{e}_{[t]}^\top\right)\left(\mathbf{H}_{[t]}-\nabla^2\mathbf{e}_{[t]}\right)^{-1}\left(\mathbf{M}^\top-\nabla\mathbf{e}_{[t]}\right)\left(\boldsymbol{\lambda}_{[t]}-\boldsymbol{\lambda}^*\right)\right\|$ in (63) is bounded.

Due to the boundedness of $\mathbf{y}_{[t]}$ and $\boldsymbol{\lambda}_{[t]}$, $\left\|\mathbf{y}^* - \mathbf{y}_{[t]}\right\|$ and $\left\|(\boldsymbol{\Lambda}_{[t]}^{-1} - (\boldsymbol{\Lambda}^*)^{-1})\mathbf{1}\right\|$ are strictly bounded away from zero. Therefore, (63) has an upper bound, which is defined as follows:

$$
\begin{aligned}
&\mathbf{B}_4 \\
&= \frac{1}{2\mu^3\boldsymbol{\lambda}_{\min}^2\{\mathbf{G}\}} \times \sup_t \Big\{ \Big\| \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) \\
&\quad \times \Big[\left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \times \Big[(\mathbf{g}_{[t]} - \mathbf{g}^*) \\
&\quad + (\nabla\mathbf{e}^* - \nabla\mathbf{e}_{[t]})\,\boldsymbol{\lambda}^*\Big] - (\mathbf{y}_{[t]} - \mathbf{y}^*)\Big]\Big\| \\
&\quad + \Big\| \left(\mathbf{M} - \nabla\mathbf{e}_{[t]}^\top\right) \left(\mathbf{H}_{[t]} - \nabla^2\mathbf{e}_{[t]}\right)^{-1} \left(\mathbf{M}^\top - \nabla\mathbf{e}_{[t]}\right) \\
&\quad \times \left(\boldsymbol{\lambda}_{[t]} - \boldsymbol{\lambda}^*\right) \Big\| + \left\|\mathbf{y}^* - \mathbf{y}_{[t]}\right\| + \left\|(\boldsymbol{\Lambda}_{[t]}^{-1} - (\boldsymbol{\Lambda}^*)^{-1})\mathbf{1}\right\| \Big\}.
\end{aligned}
$$

In summary, we have

$$
(52) \le \frac{1}{\mu}\mathbf{B}_4. \qquad (64)
$$

C. CONCLUSION

Based on the above discussion, combining (50), (60), and (64), the one-slot Lyapunov drift rate is as follows:

$$
\begin{aligned}
\Delta V\left(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}\right) &= V\left(\mathbf{y}_{[t+1]}, \boldsymbol{\lambda}_{[t+1]}\right) - V\left(\mathbf{y}_{[t]}, \boldsymbol{\lambda}_{[t]}\right) \\
&\le -\mathbf{B}_1 \left\|\mathbf{y}_{[t]} - \mathbf{y}^*\right\|^2 + \pi\mathbf{B}_2 + \frac{1}{\mu}\mathbf{B}_3 + \frac{1}{\mu}\mathbf{B}_4.
\end{aligned}
$$

REFERENCES

[1] Q.-V. Pham and W.-J. Hwang, "Network utility maximization-based congestion control over wireless networks: A survey and potential directives," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 1173–1200, 2nd Quart., 2017.

[2] J. Wen, K. Wu, C. Tellambura, and P. Fan, "Closed-form word error rate analysis for successive interference cancellation decoders," *IEEE Trans. Wireless Commun.*, vol. 17, no. 12, pp. 8256–8267, Dec. 2018.

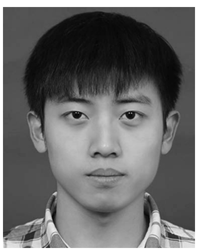
[3] J. Wen and X.-W. Chang, "On the KZ reduction," *IEEE Trans. Inf. Theory*, to be published. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8456523, doi: 10.1109/TIT.2018.2868945.

[4] A. Eryilmaz and R. Srikant, "Joint congestion control, routing, and MAC for stability and fairness in wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1514–1524, Aug. 2006.

[5] Y. Zhang, W. Xu, J. Huang, Y. Wang, T. Shu, and L. Liu, "Cross-layer optimal power control and congestion control providing energy saving for ad hoc networks," (in Chinese), *J. Softw.*, vol. 24, no. 4, pp. 900–914, Apr. 2013.

[6] X. Huang, S. Feng, and H. Zhuang, "Jointly optimal congestion control, channel allocation and power control in multi-channel wireless multihop networks," *Comput. Commun.*, vol. 34, no. 15, pp. 1848–1857, Sep. 2011.
[7] A. Zymnis, N. Trichakis, S. Boyd, and D. O'Neill, "An interior-point method for large scale network utility maximization," in *Proc. Allerton Conf. Commun., Control, Comput.*, Monticello, IL, USA, Sep. 2007, pp. 26–28.
[8] D. Bajovic *et al.*, "Newton-like method with diagonal correction for distributed optimization," *SIAM J. Optim.*, vol. 27, no. 2, pp. 1848–1857, Feb. 2017.
[9] E. Wei, A. Ozdaglar, and A. Jadbabaie, "A distributed Newton method for network utility maximization," in *Proc. 49th IEEE Conf. Decision Control (CDC)*, Atlanta, GA, USA, Dec. 2010, pp. 1816–1821.
[10] J. Liu and H. D. Sherali, "A distributed Newton's method for joint multi-hop routing and flow control: Theory and algorithm," in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2011, pp. 2489–2497.
[11] J. Liu, N. B. Shroff, C. H. Xia, and H. D. Sherali, "Joint congestion control and routing optimization: An efficient second-order distributed approach," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1404–1420, May 2016.
[12] A. Forsgren, P. E. Gill, and M. H. Wright, "Interior methods for nonlinear optimization," *SIAM Rev.*, vol. 44, no. 4, pp. 525–597, Apr. 2002.
[13] J. B. Hiriart-Urruty and C. Lemaréchal, "Convex analysis and minimization algorithms I," *Convex Anal. Minimization Algorithms*. Berlin, Germany: Springer-Verlag, 1993, pp. 150–159.
[14] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control," *IEEE/ACM Trans. Netw.*, vol. 8, no. 5, pp. 556–567, Oct. 2000.
[15] Z. I. Woźnicki, "Matrix splitting principles," *Int. J. Math. Math. Sci.*, vol. 28, no. 5, pp. 251–284, May 2001.
[16] J. Yao and Y. Liu, "Secrecy rate maximization with outage constraint in multihop relaying networks," *IEEE Commun. Lett.*, vol. 22, no. 2, pp. 304–307, Feb. 2018.

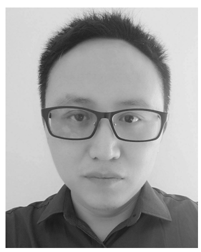
WEI FENG received the B.E. degree in electronics and information engineering from Hubei Engineering University, Xiaogan, China, in 2005, and the M.E. degree and Ph.D. degrees in communication and information systems from the South China University of Technology, Guangzhou, China, in 2009 and 2014, respectively. Previously, from 2005 to 2006, she was working as an FAE with LITE-ON Technology Cooperation, Guangzhou, China. From 2009 to 2011, she was working as a Network Engineer with Huaxin Consulting Co. Ltd., Hangzhou, China. Currently, She is currently a Lecturer with Hangzhou Dianzi University, Hangzhou, China. Her research interests emphasize on energy efficiency and physical layer security in future wireless communications, etc.



YONGXIN XU was born in Xinyang, HeNan-Henan, China, in 1997. He is currently pursuing the bachelor's degree with the School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, China. His current research interests include wireless networks optimization and distributed algorithm design, etc.

XIAORONG XU (S'09—M'11) received the B.Eng. degree in communication engineering, and the M.Eng. degree in communication and information system from Hangzhou Dianzi University (HDU), Hangzhou, China, in 2004 and 2007, respectively, and the Ph.D. degree (majoring in signal and information processing) from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2010, where he is currently with the School of Communication Engineering, Hangzhou Dianzi University (HDU), Hangzhou, China, as Associate Professor. He is also an excellent backbone teacher in HDU. From 2011 to 2013, he was working as a Postdoctoral Researcher with the Institute of Information and Communication Engineering, Zhejiang University (ZJU), Hangzhou, China. From 2013 to 2014, he served as a Research Scholar with the Electrical and Computer Engineering Department, Stevens Institute of Technology (SIT), Hoboken, NJ, USA. His research interests emphasize on green wireless SWIPT networks, cognitive radio networks (CRN), cooperative communications, energy efficiency and physical layer security in SWIPT networks, and CRN, etc.

MINGXIONG ZHAO received the B.S. degree in electrical engineering, and the Ph.D. degree in information and communication engineering from the South China University of Technology (SCUT), Guangzhou, China, in 2011 and 2016, respectively. He was a visiting Ph.D. student with the University of Minnesota (UMN), Twin Cities, MN, USA, from 2012 to 2013, and also with the Singapore University of Technology and Design (SUTD), Singapore, from 2015 to 2016, respectively. Since 2016, he has been an Assistant Professor with the School of Software, Yunnan University, Kunming, China. His current research interests are physical layer security, cooperative relay communication, and social social-aware communication systems.



YINGBIAO YAO was born in Songzi, Hubei, China, in 1976. He received the Ph.D. degree in communication and information system from Zhejiang University, Hangzhou, China, in 2006. Now, He is currently a member of the China Computer Federation (49541M), and also a Professor with the School of Communication Engineering, Hangzhou Dianzi University, Hangzhou, P.R.China. His current research interests include SSD design, wireless sensor networks and indoor localization, and multimedia signal processing, etc.